Re-conceptualizing Mental "Illness": The View from Enactivist Philosophy and Cognitive Science

Joel Parthemore and Blay Whitby (editors)

AISB Convention 2013 • University of Exeter • 3rd–5th April, 2013

Foreword from the Convention Chairs

This volume forms the proceedings of one of eight co-located symposia held at the AISB Convention 2013 that took place 3rd-5th April 2013 at the University of Exeter, UK. The convention consisted of these symposia together in four parallel tracks with five plenary talks; all papers other than the plenaries were given as talks within the symposia. This symposium-based format, which has been the standard for AISB conventions for many years, encourages collaboration and discussion among a wide variety of disciplines. Although each symposium is self contained, the convention as a whole represents a diverse array of topics from philosophy, psychology, computer science and cognitive science under the common umbrella of artificial intelligence and the simulation of behaviour.

We would like to thank the symposium organisers and their programme committees for their hard work in publicising their symposium, attracting and reviewing submissions and compiling this volume. Without these interesting, high quality symposia the convention would not be possible.

Dr Ed Keedwell & Prof. Richard Everson
AISB 2013 Convention Chairs

Published by
The Society for the Study of Artificial Intelligence and the Simulation of Behaviour
http://www.aisb.org.uk

ISBN: 978-1-908187-35-2

CONTENTS

Re-conceptualizing Mental Illness: The View from Enactive Philosophy and Cognitive Science

In the late 20th and early 21st Century, the dominant trend in philosophy of psychiatry and mental health has been toward pathologizing a wide range of mental phenomena under the headings of "disease"/"illness", "disorder", or "disability" and treating the ones labeled "illness" on a par with physical illness, to be treated primarily by drug-based interventions.

That said, certain recent trends in cognitive science and philosophy of mind – notably Andy Clark and Dave Chalmers' extended-mind hypothesis and the enactivist school associated with Evan Thompson, Francisco Varela, and others, and the Tartu school of semiotics embedding mental life into a "semiosphere" – have challenged the familiar equating of the boundaries of the physical body with those of the mind. While the various approaches differ at key points, all agree that, although the mind must be physically realized, it extends in substantive ways into the environment, its boundaries subject to constant negotiation and re-negotiation.

As such extended-mind critics as Robert Rupert point out, re-conceptualizing the boundaries of mind and world in this way can only be justified if there is some empirical payoff. A small but increasingly vocal group within the extended-mind/enactive community believe that one of the best places to look for such payoff is in the field of mental health. They suggest moving away from a model based on physical illness towards one that emphasizes each person's history and embedding in a social context: such identified conditions as Asperger Syndrome and high-functioning autism may be better understood as instances of cognitive diversity rather than impairment; while conditions such as schizophrenia or manic-depressive disorder must be understood, and treated, as problems of the patient's immediate community and not just the patient herself. Furthermore, they must be understood, and treated, in light of the patient's history of interactions with her environment and not just the presenting symptoms. The risk of much contemporary treatment is that, like aspirin, it treats the symptoms and does not address the underlying issues.

As an emerging community and not just a scattered collection of "lone voices", the field is brand new, and it is cutting edge. It touches on such key themes as the nature of mind and its relationship to environment; the possibilities for computer models of mind that draw on exciting new paradigms; and the breadth of cognitive science, from theoretical explorations in philosophy of mind to concrete applications and new directions in treatment.

Seven of the nine papers included in the present volume address these issues from various directions: from philosophy of mind, psychology, psychiatry, and cognitive science; from broad overview or specific diagnosis; from more critical and more sympathetic perspectives on enactivism. The final two papers place the discussions into a wider context of enactivism: its foundations and its future.

Along with these contributors – and a tenth contributor, Sanneke de Haan (University of Amsterdam), whose contribution is not able to appear here because of its submission elsewhere – we welcome Mark McKergow of the Centre for Solutions Focus at Work (http://www.sfwork.com) and Nick Medford of the Brighton and Sussex Medical School as our keynote speakers.

Chairs

Joel Parthemore (Centre for Cognitive Semiotics, University of Lund, Sweden)

Blay Whitby (Centre for Research in Cognitive Science, University of Sussex, UK)

Program Committee

Fred Cummins (School of Computer Science and Informatics, University College Dublin)

Sanneke de Haan (Department of Psychiatry, University of Amsterdam)

Dan Hutto (Professor of Philosophical Psychology, University of Hertfordshire, UK)

Marek McGann (Department of Psychology, Mary Immaculate College, University of Limerick)

Göran Sonesson (Centre for Cognitive Semiotics, University of Lund, Sweden)

No Mental; Health

Marek McGann[1] and Fred Cummins[2]

Abstract. The basic argument to be made has two parts. Part the First: there is no tenable distinction to be made between the mental and the physical. The enactive approach is probably the best framework for expressing this constraint among contemporary theories. Part the Second: As we move from consideration of the identity of cells to the identity of nations, there is no single level that is co-extensive with the person. In particular, identification of the person with the human body is inappropriate. The enactive approach is ambivalent here, and recent attempts to provide definitions of agency seem to run the risk of fixing the person, inappropriately, at one level or another. The consequences of these two observations is that there is no coherent domain of mental health. There is health: the health of cells, of bodies, of families, of football teams, and of nations.

1 INTRODUCTION

The field of mental health lives a precarious existence. To those of a neuroreductionist bent it is just another medical speciality, on the same level as immunology, gynaecology and oncology, and disorders of the mind are nothing more or less than disorders of the brain. Those who favour a functionalist, cognitive psychological account will see mental disorders as malfunctions of a notional cognitive system, for which there are assumed to be norms that allow distinctions between healthy and pathological operation.

The day to day life of the practitioner in the field of psychiatry forces a somewhat more eclectic and pragmatic view, as cases that present display a very wide range of problems that need to be confronted. Some are clearly of organic origin, as in frontal lobe tumours; some may have organic correlates, such as serotonin imbalance in depression, but the problematic manifestations frequently lie rather in the lived experience of the subject. Making a link from the observed problematic to the presumed level of physiological regulation becomes increasingly more difficult as we move among the cornucopia of neurotic and psychotic phenomena, many of which display no obvious link to physiological disorder at all, and many of which may plausibly be argued to be disorders of a social field, rather than a single biological individual [1], or to reflect culturally specific normative considerations unrelated to the body [2]. Some cases that present may appear to be problems of behaviour or belief with no identifiable organic pathology. Even if one were fully subscribed to either a neuroreductionist or a functionalist interpretation of the field, there is little hope that such accounts will make significant contributions to many of the problems faced by clinicians in psychiatry in the short to medium term.

In this brief contribution we seek to provide a basis for an alternative discussion of such issues. The argument to be made has two parts. Firstly, we argue that there is no tenable distinction between the mental and the physical. Adopting this stance will affect how we frame all subsequent discussion of mental and physical health. We believe that the enactive framework that is emerging may be the best of the current stock of theoretical approaches to develop arguments that are free of the mental-physical dichotomy. We then follow a line of argumentation that is frequently followed in the enactive literature, among others, to consider the relations that obtain among levels of systemic organisation in living beings, from the cell to societies. Here, there is much work yet to be done, and we will argue that there is no level in this hierarchy (or, better, network) that is co-extensive with the person. This second claim runs counter to some recent proposals within the enactive literature about the nature of agency, and we suggest that there is an important discussion waiting to happen here.

2 THE STICKY LEGACY OF MIND-BODY DUALISM

Contemporary understanding of health and well-being remains strongly affected by the legacy distinction between the mental and the physical. To the neuroreductionist or eliminativist, the need to treat the mental as *sui generis* is a pragmatic step, necessary for the daily conduct of business until the job of translating the vocabulary of the mental into the vocabulary of neural events is complete. We are not holding our breath.

To the functionalist, or cognitivist, mental health issues may be presumed to pertain to the (dys)function of a notional cognitive mechanism that is most properly characterized on its own terms as information processing and computation over representations of the world. Difficulty arises as measurement and observation are strictly limited to features of the world, and the presumed cognitive machinations are never directly observable.

In both approaches, and many similar discussions, most of the considerable energy spent has been directed towards trying to shoehorn a problematic domain of the mental into the domain of the physical, presumed to be somehow simpler, or less in need of justification. To the eliminativist, the physical is the only real level of description; to the functionalist, it is the only observable domain.

There seem to be two senses of the term physical that are

1 Dept. of Psychology, Mary Immaculate College ~ University of Limerick, South Circular Road, Limerick, Ireland. Email: marek.mcgann@mic.ul.ie
2 School of Computer Science and Informatics, University College Dublin, Belfield, Dublin 4, Ireland. Email: Fred.cummins@ucd.ie

lurking in the background here, and when examined, neither suggests that the physical is any more easy to pin down than the mental. These two senses we will call the 'experiential' and the 'theoretical', though inevitably, hybrid accounts can be found.

The experiential-physical is best illustrated by the search of Doubting Thomas for the kind of proof that is beyond question. Thomas insisted on seeing the risen Christ with his own eyes, and placing his own finger in the wounds. A similar appeal to the experience of a reassuring presence is illustrated by Doctor Johnson who famously sought to repudiate Berkeleyian Idealism by kicking a stone and saying "I refute it *thus*". This is the sense of physical indubitableness that arises in our everyday experience of tangible objects and substances. But of course, for Doubting Thomas and Dr Johnson alike, the satisfaction of physicality arises only through sensorimotor engagement with the world, through visual or haptic exploration on the one hand, and through the wilful act of kicking on the other. Both require feats of perceptually guided action and both provide the security blanket of the ``physically real" only in the direct experience, or sense-making activity, of the doubter. This sense of physical is thus not clearly or conceptually distinct from the mental at all.

But perhaps the examples cited above relate only to bar-room argumentation, and the term ``physical", when employed in scientific debate actually means something rather different, viz. that which is the object of the science of physics. And then we must ask, on which physical theory do we hang our hat? For physical theories no longer trade in the substantial and tangible, but in strings, quarks, and fields, all of which are very far removed from the indubitable, tangible, and kickable. Newtonian physics may no longer be state of the art but it is physics that was developed to account for the motions of massive objects at spatial and timescales that were familiar to scientists. It is the best physics in the world to describe the carry on of apples and missiles, and its very effectiveness for such objects is testament to the deep link between the framework and tools of the Newtonian enterprise and the embodied reality encountered by organic beings of a specific size and with a specific metabolic rate. But these characteristics do not permit the separation of a ground for reality from the experiences of conscious embodied beings; rather, they emphasise the very deep interconnection between the world contingently experienced by such beings, and the theory that best accounts for measurement in such a world. And they fail to make any meaningful link whatsoever to modern physics, where measurements are made at spatial and time scales vastly larger or smaller than those centred in the body.

This failure to identify the ``physical" undercuts any attempt to use such an identification to then characterize some notional ``mental" domain. If we accept this, then eliminativism, or neuroreductionism, becomes incoherent. Absent a mental-physical distinction, there are no distinct concepts to be reduced to neural events.

Our goal here is not to do metaphysics but to seek a path forward, beyond the intellectual traps that continue to license the inclusion of phenomena in the ontological dustbin of "mental health". Many have sought to abandon the mental-physical distinction, without reducing the level of explanation to one or the other domain [3]–[5]. Within contemporary approaches to cognition, the Enactive approach, Ecological Psychology, Coordination Dynamics, and recent initiative such as Radical Embodied Cognitive Science [6] or Radical Embodied Cognition [7] all adopt a vocabulary that nowhere acknowledges or relies on a mental-physical distinction. But it is within the mind and life or enactive approach that is emerging from foundations in the work of Varela, Thompson and others, that this dualism is rejected in a most principled way. In an enactive account, no mental-physical distinction arises (Varela, Thompson, & Rosch, 1991). The bringing forth of a world, as Varela called it, through the sense-making activity of an autonomous system at once acknowledges the dual subjective/objective character of the lived world, and sidesteps most of the legacy mind-body dualisms inherent in received approaches. There thus seems to be some *prima facie* reason to believe that problems that have proven intractable within paradigms irrevocably committed to a subject-object, or mind-world, distinction may be addressed in a new light within an enactive framework. If the concept of ``mind" does not stand in opposition to the concept of ``body", then there is little justification for distinguishing between ``mental" and ``physical" health.

3 MANY MODES OF DESCRIPTION

Biological agency, particularly human agency with which we are concerned here, has we might say, many moving parts. Agency arises within a very dynamic and complex web of phenomena, from the cellular to the cultural. Biological requirements and appetites wax and wane within socially structured opportunities and physical affordances that enable, invite and scaffold actions.

If we use an efficient causal framework – one cause to one effect in billiard ball-style progression – then human behaviour is a compromise formed within the interaction of many different biological factors, conditioned within multiple physical, social and developmental constraints.

This matrix of causality resists explanation in terms of a single canonical thread running forward through time, identifiable as “the agent” or “the person”. We do not argue here that the explanation of action at any given time is arbitrary. However the perspectival nature of any explanation, including those from a first person point of view, means that there is no single correct explanation that captures the totality of the person at a given moment in time and exhaustively explains their behaviour.

The descriptions of behaviour that we choose to offer at a particular time are characterised by that subset of the field of forces within which the person we are describing is moving, that we ourselves can perceive, and that enable us to make sense of, to sensibly coordinate our own behaviour with, as best as possible at that time.

It is part of the ambition of enactive theorists to do more than offer neutral descriptions of actions, however [8]–[10]. The

enactive literature is replete with discussion of value and normativity, the perspective defining character of mindedness on which is (or will be) founded a satisfactory understanding of meaning and experience [8], [11]–[14]. Despite the prevalence of such discussion however, there remains some confusion as to what might be the ultimate ground for value, or how different values might be related. Two themes can be discerned, one on the foundational role of operational closure and autopoiesis as the biological fundament of meaning and mind (biological autonomy), while other work focuses on the more fluid autonomy of social interactions (participatory sense-making).

Systemic Value, Growth and Identity

It is a core tenet of the enactive approach that values for a given system arise inherently within the operation of the system itself. Values, norms and normativity are not determined by the comparison of current state with an ideal (or even "normal") state but are enacted in the operation of the system itself over time [8], [12].

The most basic systemic value is that of identity grounded in organisational closure. The concept of identity in play here is somewhat slippery. Its clearest definition is derived from Varela's [15] work on biological autonomy. Relations of mutual support or dependence between components of a system form an implicit identity – the circularity of the relations effectively instantiating a distinction between the system and its surround, the production-in-action of an identity for that closed network of components (see Figure 1.).

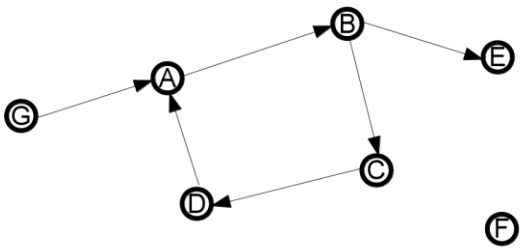

Figure 1. An organisationally closed network of production, ABCD, forms a dynamic identity. Nodes are processes of production (usually considered as biochemical processes) arrows are relations of support. Because they are not mutually supporting within the network, nodes E and G are not part of the autonomous system per se.

It is standard discourse within the enactive literature to hold that such dynamically constituted identities form the basis of normative, value-driven activity, the foundation of agency (Thompson & Stapleton, 2009; Weber & Varela, 2002). The value of self-maintenance is inherent in the system because of the manner in which the system's organisation operates so as to maintain itself. Should the organization break down the identity is lost – this is something basic and intrinsic to the system itself, not something that depends on the observation of a third party or the judgement of a dedicated subsystem of sensors and comparators (Di Paolo, De Jaegher & Rohde, 2010).

Though it is certainly not offered as the full story (see particularly Di Paolo, 2005, also Di Paolo 2009 and Barandiaran, Di Paolo & Rohde, 2009) it is from such a concept of identity and value that the enactive literature to date promises to adequately address questions of agency, meaning and normativity. The logic of organisational closure, systemic value and dynamic identity is sufficiently generic that it "scales up" and can be applied in some way to social and personal forms of identity (see De Jaegher & Di Paolo, 2007; Di Paolo, 2009). Whereas in the biological case the components of the network are biochemical processes of production, in the social case the nature of the components is less clear. Candidates include habits, cultural practices or skills structured by sensorimotor contingencies (or some personal, emotional equivalent; see McGann & De Jaegher [16]). The identification of any set of components, and the identification of any superordinate systemic domain may depend, *inter alia*, on the purposes of the investigator.

The autonomy provided by organizational closure is seen as fundamental to agency and the agent as effectively co-extensive with the identity of the system in question. The domain of cognition is to be interpreted as the domain of relations between the agent as identified in the network of self-production and its environment.

Our problem is that it is difficult to imagine only one such agent, and only one identity as defined here, existing in the complex of relations that encompass a human life. Rather than a single identifiable entity around which sets of normative relations might be sought, it is likely that a skein of such identities might be available to astute observers, with no single identity (or its attendant systemic value) having any cause for claims of precedence.

We argue that there are, in any given action, values inherent in the action that are produced by the organisation of the tangle of forces in the field that define that situation (its biological, social, developmental and "merely" physical facets). None of these values can be identified as the ground, the ultimate foundation from which a detailed and comprehensive account of the action in question must be built. Making sense of any moderately complex human behaviour necessitates appeal to many domains of organization simultaneously. The biological domain of the body is one such, but it is neither the sole, nor often the most important domain within which behaviour is grounded.

Life is dynamic, and while at one moment I might act such as to ensure a particular value is maintained, the fine details of how that value is instantiated and how it influences my behaviour over even short timescales may vary as the situation changes. There will be no one moment where the "true" intention of an action can be identified, just a prolonged period of time over which the action can be understood if the actions of the perceiver can be brought into coordination with it in some way.

Rather than self-maintenance of any kind, we must seek to

ground our ideas of value in continuity. If we imagine Figure 1. as indicating personal, social or cultural identities being formed in the dynamic of cultural practice and social interaction it is easy for us to imagine, as the situation evolves, as the relationship between the interactants changes (perhaps just over the order of seconds), that the relationships between the components of interest to us, the observers, changes.

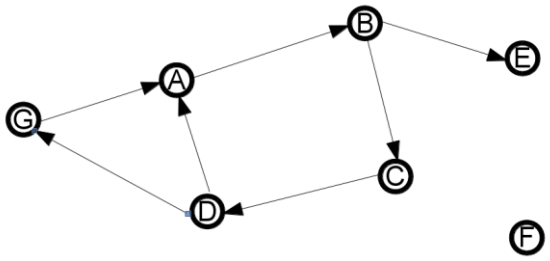

Figure 2. An evolving identity. The initial identity might be subsumed or in a later step even destroyed, but the agent, the relations of interest show an important continuity that does not challenge the existence of the agent of interest to us (and maybe not to the person themselves either).

There is no privileged level of description, no privileged vocabulary of description and no single canonical agent or even canonical action.

It has been an continuing implication of enactive thinking that wherever cognitive scientists seek to find a clear distinction, some solid ground on which to stand a theoretical edifice, that the approach breaks that ground into shifting sands. What is needed is not a firm place to stand but a willingness and ability to move with the dynamic, in-flux, phenomena that we seek to describe. The concept of health will be no different here.

4 HEALTH AND SYSTEMIC VALUES

Barandiaran and Egbert [12] have outlined a mathematical framework for trying to capture the idea of a momentary, systemic value. They endorse an organisational view of normativity – that a system can by virtue of operational closure, instantiate a norm of self-maintenance. They also point out, however, that such norms are "virtual". They are not mechanisms or special-purpose components of a given system. They are instead emergent phenomena within the space of possible actions of the system, some of which will lead to its survival, some to its death, but which must be understood and evaluated dynamically, moment to moment. Their virtuality is a product of their dynamism.

Their paper presents their framework for the minimal case of a single cellular entity engaging in chemotaxis. They suggest (perhaps at this point rather optimistically) that, where there are multiple values instantiated by a given entity, that the method's principles of analysis will remain. A very great deal of work remains to identify whether this promise is possible, let alone to actually follow through with such developments, but their analysis at least provides us with one way of considering the concept of health within an enactive framework.

Health is an expression of a system's values. If a system, howsoever described, can be said to be operating in such a way as to continue to operate and (if appropriate) achieve its ends, then it is healthy.

Ill-health then is a relational characteristic that describes the behaviour of a system as inviting intervention. The perception or experience of ill-health is an expression of the values of that system, be they biological, personal, social, cultural or otherwise.

We see problems when various forms of these values fall into conflict with one another – personal experiences of normal behaviour may conflict with cultural norms, for instance, or with biological, such that the larger system is unhealthy while the elements comprising it could be considered perfectly fine. By contrast we might see perfectly healthy systems whose components are run down and destroyed by its "healthy" operation. The health of one system may indeed represent a threat to the health of another, as in the conflict between a tumour and its host, or a paramilitary force and a nation state.

In real terms, this is a more general statement of the concept of defining health less as the "normal" operation of a biological or psychological system and more as a matter of quality of life[17], [18]. We see the logic of enactivism, brought to its own conclusions, as providing a principled means by which such contextualised, observer-dependent judgements of health, quality of life can be stated, and by which the continuity of such considerations across the range of human experience can be framed.

Given the fragmented, piecemeal nature of much of the discourse concerning mental illness and psychological disorder at present we consider this a valuable contribution of an emerging enactive framework. The real work of developing from this beginning a fully development paradigm of health and living still proves a mammoth, if inviting, challenge.

5 CONCLUSION

The ideas, diagnoses and judgements of ill-health must be made with an explicit framing of the context in which those judgements are being made, and the values against which observed behaviour is seen as inviting intervention. Practitioners must be sensitive to the potential for conflicts between different modes of description of behaviour, including the system in which both they and their clients are mutually influencing components. Questions of health are not independent from questions of systemic values and the shifting boundaries of system identity, where the systems in question may range from the sub-cellular to the societal. There is thus no domain of mental health. There are questions of health, period.

REFERENCES

[1] R. D. Laing, *The Politics of Experience : And, The Bird of Paradise*, First Edition. Penguin, 1967.
[2] L. J. Kirmayer, R. Lemelson, and M. Barad, *Understanding trauma: integrating biological, clinical, and cultural perspectives*. Cambridge: Cambridge University Press, 2007.
[3] L. Wittgenstein, *Philosophical investigations*, 2d ed. Oxford: Blackwell, 1958.
[4] G. Ryle, *The Concept of Mind*. Oxford: Oxford University Press, 1949.
[5] M. Merleau-Ponty, *The Phenomenology of Perception*. London: Routledge & Kegan Paul, 1962.
[6] A. Chemero, *Radical Embodied Cognitive Science*. Cambridge, Mass: MIT Press, 2009.
[7] D. Hutto and E. Myin, *Radicalizing Enactivism: Basic Minds without Content*. Cambridge, Mass.: MIT Press, 2013.
[8] E. A. Di Paolo, M. Rohde, and H. DeJaegher, 'Horizons for the enactive mind: values, social interaction and play.', in *Enaction: Towards a new paradigm of cognitive science*, J. Stewart, O. Gapenne, and E. Di Paolo, Eds. Cambridge, Mass: MIT Press, 2010.
[9] M. McGann, H. De Jaegher, and E. Di Paolo, 'Enaction and Psychology', *Review of General Psychology*, vol. 17, no. 2, in press.
[10] A. Weber and F. J. Varela, 'Life after Kant: Natural purposes and the autopoietic foundations of biological individuality', *Phenomenology and the Cognitive Sciences*, vol. 1, no. 2, pp. 97–125, Jun. 2002.
[11] X. Barandiaran, E. Di Paolo, and M. Rohde, 'Defining agency:individuality, normativity, asymmetry and spatio-temporality in action.', *Adaptive Behavior*, vol. 17, no. 4, pp. 1–13, 2009.
[12] X. E. Barandiaran and M. D. Egbert, 'Norm-establishing and norm-following in autonomous agency', in press, *Artificial Life*.
[13] E. Di Paolo, 'Autopoiesis, Adaptivity, Teleology, Agency', *Phenomenology and the Cognitive Sciences*, vol. 4, no. 4, pp. 429–452, Dec. 2005.
[14] E. Thompson, *Mind in Life Biology, Phenomenology and the Sciences of Mind*, 1st ed. Harvard University Press, 2007.
[15] F. J. Varela, *Principles of Biological Autonomy*. Norwalk, CT: Appleton & Lange, 1979.
[16] M. McGann and H. De Jaegher, 'Self–other contingencies: Enacting social perception', *Phenomenology and the Cognitive Sciences*, vol. 8, no. 4, pp. 417–437, Dec. 2009.
[17] D. B. Double, *Critical psychiatry : the limits of madness*. Basingstoke: Palgrave Macmillan, 2006.
[18] P. M. Rapley, D. J. Moncrieff, and M. J. Dillon, Eds., *De-Medicalizing Misery: Psychiatry, Psychology and the Human Condition*. London: Palgrave Macmillan, 2011.

Isolated Sailors in Isolated Ships: the Case of Autism

Anna CIAUNICA[1]

Abstract. It has become fashionable to think of limitations in psychological perspective-taking among children with autism as signs of lacks in their "Mindreading" abilities [1,2]. On the dominant view, it has been argued that the human aptitude at inferring mental states is one of the crucial preconditions for the evolution of cooperative social structure in human societies. In this paper I argue that, on the contrary, modes of social relatedness may be constitutively primary over modes of cognitive abilities. In a first step I argue that standard "mindreading" theories may have overestimated the role of conceptual thinking and underestimated the role of intrinsic social-emotional organization. Then I review behavioral and neural evidence illustrating that in autistic individuals, an impairment of the relational self could trigger a shortfall of both (i) the basic portfolio of cognitive abilities and (ii) the coordination of intentional activity in social interaction. Finally, this paper supports the conclusion that what is missing among autistic children highlights what is present among children without autism, namely forms of emotional *engagement* through which a child is moved in psychological attitudes by the bodily expressed attitudes of someone else [3].

1 INTRODUCTION

Central to the study of the development of self and self-awareness are questions that concern relationships between a self and other people, and a self and the physical world. The issue at stake is crucial since through the approach of developmental psychology – and specifically the study of autism[3] – we may find structures through which "primitive" forms of interpersonal links are achieved. According to a longstanding philosophical tradition, to know is to have a mind that process detailed internal representations of the outside world. It is the philosopher's job to accurately reproduce such a mind's representing powers and to seek for an unshakeable foundation, immune from all doubt, i.e. the "firm basis on which the tottering edifice of our knowledge is reared" [4].Traditionally, it has been argued that the human aptitude at mentalizing is one of the crucial preconditions for the evolution of cooperative social structure in human societies. Hence, lack of mentalizing abilities might induce social connectedness impairments. Indeed, in Autistic Children (AC henceforth), for reasons that are not yet understood, there are failures of both the basic portfolio of mentalizing skills and at the interpersonal connectedness level. It has been hypothesized that these children have a specific *Mindblindness* deficit [1]. Both clinical descriptions [5] and experimental studies [6, 7] illustrate how AC appear to lack a grasp of self and other in social interactions. Indeed, systematic investigations have revealed that AC show: (a) limited responsiveness to others in settings that elicit social referencing [8, 9]; (b) impairment in the assimilations of the stance of the other [10, 11]; and (c) reduced one-to-one intersubjective engagement and responsiveness along with impairment in joint attention and other forms of 'secondary intersubjectivity' [12]. Possible mindblindness explanations about the primary causes of autism typically include: i) a missing drive for global coherence [13]; ii) a missing theory of mind module [1,2]; iii) a deficient eye tracking module[14]; iv) a deficient attention switching device [15]; v) an executive function deficit [16]; vi) a deficient imitation mechanism [17]; vii) a deficit of the contingency detection module [18].

Over the past two decades there has been a two-party debate between Theory Theory (TT) and Simulation Theory (ST). TT theorists [1, 19, 20, 21] hold that we explain and predict others' behavior by relying on (i) an innate modular or (ii) acquired mechanism/theory of how people generally behave. ST theorists [22] argue that we explain and predict others' behavior by using an inner model designed to simulate other person's mental states. Note that both TT and ST accounts resort to third person-based explanation and prediction in order to understand ordinary social interaction. Within the last ten years, a new party to the TT/ST debate has emerged, i.e. the embodied/enactive/situated approach [23, 24, 25, 26, 27, 28], challenging the Mindreading thesis. These theorists reject the idea that we understand others by having some grasp of their mental states and using this understanding to explain and predict their behavior on the basis of these mental states. They also deny that at the core of our intersubjective skills there are high-level cognitive mechanisms. The new Embodied Social Cognition (ESC) approach [29] is roughly the view that our normal everyday interactions consist in non-mentalistic embodied engagements. Since so much of my line of reasoning hangs on the contrast between Mindblindness and ESC views, it is worth going over familiar ground carefully in preparation for the idea that children with autism lack metarepresentational abilities *because* their relational self is impaired, and not the other way around

2 THE DEVELOPMENTAL CLAIM AND THE "STARTING PROBLEM"

Within the standard interpretation, there is a well documented debate between *nativists* and *empiricists*. The nativists argue that there is an innate Theory of Mind Mechanism (ToMM) [20]. By contrast, the empiricists argue that children learn about mental states by putting at test different rival theories about the behavior of other in social interaction [30]. A full review of this literature is out of the question here, given the space available. Instead I will attempt to expose in rough detail the most fundamental problems facing those who hope to explain social cognition and

[1] Dept. of Cognitive Science, Univ. of Lyon (France) & Dept. of Philosophy, Univ. of Fribourg (Switzerland), av. de l'Europe – 1700 Fribourg, Email: annamaria.ciaunica-garrouty@unifr.ch.

[3] Autism is a developmental disorder with a genetic basis and a prevalence of 0.1 to 0.6%. It is diagnosed on the basis of early emerging qualitative abnormalities in social interaction, communication and imagination. Given the limited space available here, I shall use the term "autism" as an umbrella term for "Autistic Spectrum Disorders".

autistic impairments on solely mentalizing grounds. For our discussion here, it is important to outline that both nativists and empiricists share what Spaulding calls the *developmental claim* (DC):

> $DC_{Mindreading}$: a developed capacity for social understanding and interaction requires that children (*somehow*) come to understand mental states [31:121,emphasis added].

Clearly, a lot of explanatory weight rests on the shoulders of the word "somehow" sneaked between parentheses in the quotation above. The idea is to examine closer the interpretation of this mysterious "somehow" and illustrate that TM/ST do not provide a satisfactory solution to what Gallagher [32] recently coined as the "starting problem". The "starting problem" is the challenge to explain "how the social cognitive process gets off the ground" – or more precisely what ground we stand on as we engage in the process. Before I start, I hasten to add that my summary of this literature should by no means be viewed as exhaustive. A classic debate in developmental literature concerns whether infants are initially more sensitive to internal or to external stimuli. On the one hand, it has been hypothesized that infants make use of innate cross-modal capacity to map the adult's visual movements onto the proprioceptive feelings of his or her own movements that the adult is imitating. For example, Meltzoff and Gopnik [17] proposed that there are innate mechanisms that allow the infant to attribute emotions to other minds starting from birth. Experiments on neonatal imitation [33, 34] and on the innate basis for primary emotions [35, 36, 37, 38] suggest that by imitating the caregiver's facial emotion expression, the infant activates through prewired connections [36] the corresponding physiological emotion state in himself or herself. The imitation-generated internal emotion state is then introspectively accessed, and the felt affect is attributed to the other's mind. The basic assumption behind this view is that the infant's initial state is characterized by direct introspective access[4] to internal emotion states, and that they have conscious access to their internal basic emotion states from the beginning of life. By contrast, Gergely & Watson [18] disagreed with this view and argued that at the beginning of life "the perceptual system is set with a bias to attend to and explore the external world and builds representations primarily on the basis of exteroceptive stimuli." In other words, infants are born with an innate Contingency Detection Module. These debates have been abundantly discussed in the literature and the details will not be pursued here. For the moment, suffice it to say that the general idea is that "on the mindreading picture, *once* the child comes to know about mental states, he or she learns how to attribute them and then how to exploit this ability in order to explain and predict behavior on the basis of mental states. How this works is the subject of the debate between the Theory Theory and the Simulation Theory" [31, emphasis added]. There is an important worry with this interpretation since the mindreading accounts provide explanations about how children attribute mental states *once* they already come to know about mental states, but it does not address the crucial question of explaining *how* exactly this is supposed to happen. Here is an example of an attempt to sweep the explanandum under the carpet without really explaining it: "First, the attributor creates in herself pretend states intended to match those of the target. In other words, the attributor attempts to put herself in the target's 'mental shoes'" [43]. But as Gallagher [32] pointed out, the question is: how do I know which pretend belief state matches what the other person has in mind? Isn't this what simulation is supposed to deliver? If I already know what state matches the target, then the problem, as defined by ST, is already solved. Likewise, TT defenders face the same difficulty. Recall that according to TT we are phylogenetically designed and innately predisposed to attend to mental states over other things. But even if we are innately tuned to attend to mental states, how do we know *what* those mental states are in any particular case? Hence the "starting problem" poses a threat to any solution that appeal to internal mechanisms.

3 COUNTERMIRROR AND SENSORIMOTOR LEARNING

There is now active research into the relations between the mirror neurons[5] and social cognitive processes such as understanding the intentions of other people. It has been hypothesized that mirror neurons, by mapping goal-directed motor acts, allow a *direct* form of action-understanding through a mechanism of "embodied simulation" [45, 46]. This challenges traditional mindreading theories share the idea that social understanding is a matter of projection of inner representations onto others. Rather at the basis of a capacity to understand others' intentional behavior, there is a more basic functional mechanism which exploits the intrinsic functional organization of the parieto-motor circuits like those containing mirror-neurons. More specifically, the motor system, by anchoring the multimodal integration, enables social connectedness. Thus, it has been claimed that mirror neurons provide a new empirical-based notion of intersubjectivity viewed first and foremost as "intercorporeity" [46], i.e. the *mutual resonance* of intentionally meaningful sensory-motor behaviors, which is the main source of knowledge we directly gather about others. By activating the neural systems underpinning what infants and agents do and feel, they obtain a direct form of understanding of others from within, as it were. Consequently, the hypothesis that dysfunction of mirror neurons might be one of the core deficits of socially isolating disorders such as autism is now under active scrutiny [47, 48]. A full review of this literature is beyond the scope of this paper. At this point of the discussion, what seems to be important is that recent evidence suggests that sensorimotor learning *can reconfigure* the mirror system. In an ingenious experiment, Catmur et al. [49] showed that human mirror system is, to some extent, both a product and a process of social interaction and that the development of the mirror system depends on sensorimotor learning. They measured mirror-system functioning before and after incompatible ("countermirror")

[4] Bruner et al. [39] also proposed that the infant moves from an initial reliance on internal, proprioceptive cues to a reliance on exteroceptive cues. See also [40, 41].

[5] Mirror neurons have been identified in two cortical areas — the posterior part of the inferior frontal cortex and the anterior part of the inferior parietal lobule. Crucially, the anatomical location of the mirror neuron system is a key feature to understand the nature of its functions. Its proximity to frontoparietal systems that support various forms of sensorimotor integration suggests that the nature of action coding implemented by the mirror neuron system is also linked with some form of sensorimotor integration [44].

sensorimotor training, in which human participants performed index-finger movements while observing little-finger movements and vice versa and provide strong support for the theory that the "the mirror properties of the mirror system are genuine but *not intrinsic*—they depend on the experienced contingency, rather than the objective similarity, between stimuli and responses" [49:1529, emphasis added]. In other words, if an infant were unable to see her own actions and grew up in an environment where mirrors and imitating adults were replaced by systems that showed counter-mirror actions (e.g. foot movements when she moved her hands), then she would develop a counter-mirror system. This encourages us to replace the above mentioned $DC_{Mindreading}$ by the following claim:

> $DC_{Enaction}$: mentality-constituting interactions are grounded in, shaped by, and explained by nothing more than the history of an organism's previous interactions [50:8]

In this section I argued that if the mirror system is *not* structured via intrinsic and primitive building blocks ready to be detected in the brain. The mirror-system responses develop as a result of general processes of sensorimotor associative learning. Given that social interaction is an important source of the sensorimotor experience, then the later is necessary for the configuration of the human mirror system. In the next section I argue that attempts to find specialized cognitive modules in individuals' heads might be misguided. Building upon [51] distinction between the *interpersonal* and the *ecological* self and starting with premise that the early manifestation of a sense of self in the physical and social domain is not a given but rather "develops via the active process of intermodal perception and exploration" [52:515], the idea is to argue that that failures in emotional engagement and perspective-taking may be *primary* in development and *cause* autistic children's cognitive limitations.

4 THE RELATIONAL SELF IMPAIRMENT

Mindreading theorists such as Lombardo & Baron-Cohen [53] recently acknowledged that all proposed mechanisms for mindblindness in AC have been very "other"-centric in nature focusing on how individuals read social cues from others (e.g., facial expressions, eye gaze, body postures), or have been agnostic with respect to the target of mentalizing (e.g., intentionality detection, mental state representation). Hence these researches have left a *gap* in terms of mechanisms that may be responsible for atypical self-referential processes in AC and their integration into the bigger picture of how individuals navigate and interact with the social world. What is needed is a more *relational* approach in order to reveal the deeper complexities involved in interpersonal relations. While I agree with Lombardo & Baron-Cohen on this final point, I disagree with their proposal to replace the search of *individual devices* with a search of *relations between devices*. Rather the social relatedness itself might be constitutively fundamental to metalizing abilities. A good way to bridge the above mentioned gap is to examine whether the aspects of social-cognitive impairments that extend beyond the autistic children's limitations to form representations are foundational in nature.

Against theories focusing upon cognitive aspects of autism, Hobson [3, 7, 10] has long championed a view which insists on the idea of a profound disruption of patterned intersubjective engagement between the child and other is basic in autism. *An impairment of the relational self (RSI)* is typically characterized by a dissociation between autistic children's self-consciousness in being observed and their ability to be *affected by* and *engaged with* the attitudes of a particular embodied person. Indeed, a deficit at the level of interpersonal connectedness could prevent AC to register and assimilate the bodily-anchored psychological stance of another person as suggested by a recent study [54]: AC witnessed one adult tearing another (nonresponsive) adult's drawing. In contrast to children without autism, who expressed dismay, questioned the perpetrator, and showed concern towards the victim, most children with autism showed very little indication of *feeling for* the person whose drawing it was. The victim had shown no overt expression, yet the children without autism immediately orientated towards, and showed concern for, this person [61].

A useful conceptual tool in examining these *affective engagement deficits* might be the notion of "participatory sense-making" as championed by De Jaegher & Di Paolo [29]. In a nutshell, "participatory sense making" is the process of generating and transforming *meaning* in the interplay between interacting individuals and the interaction process itself. Indeed, in typically developing individuals, coordination affects individual sense-making because they are constantly and directly *affected* by the coordination of movements in interaction. Similarly, fMRI evidence is accumulating that infants' brain organization in typically developing children may well adapted to be an "intersubjective system" [55, 56]. Furthermore, developmental studies seem to suggest that higher level cognitive processes are strongly affected by the history of social interactions [57, 11, 58]. The next section reviews behavioral and neural evidence supporting the RSI.

5 BEHAVIORAL AND NEURAL EVIDENCE

If we admit that sense-making is a relational and affect-laden process grounded in biological organization then an impairment of the relational self could trigger a shortfall of the participatory sense-making process, i.e. of the coordination of intentional activity in interaction. The RSI would be consistent then with several *behavioral* and *neural* findings about AC.

I) At the *behavioral* level it has been pointed out that while normal and mentally handicapped children recall meaningful sentences better that random word strings, AC are almost as good as recalling the later as the former [59, 60]. Also AC make less use of context and pay preferential attention to parts rather that wholes. Furthermore, Klin et al. [62] found deficits in early social skills such as reaching in expectation of being picked up: these impairments in primitive social abilities are not easily explained as the result of impaired cognitive abilities. Typically, the litmus false-belief test in mindreading literature [63] uses elicited-response task in which children answer a direct question about an agent's false belief. However, studies using spontaneous-response tasks such as the violation-of-expectation (VOE)[6] paradigm [64] illustrated that AC failed in this false-belief spontaneous attribution [65]. This indicates that their impairment in false-belief attribution is *independent* of their verbal abilities. Mindblindness defenders explained this failure

[6] VOE task tests whether children look longer when agents act in a manner that is inconsistent with their false beliefs.

by arguing that AC display a fundamentally cognitive impairment in ToM. However, Turner [66] (1996) illustrated that the degree of repetitive behavior in autism is not related to ToM abilities nor to IQ. Indeed, some types of repetitive behavior (tics and motor problems) were more severe in those with high verbal abilities, many of whom pass the false belief test. Hence, a much more plausible explanation is that AC lack *social motivation* or *engagement* for tracking the agent's epistemic state. For example, Ruffman et al. [67] also found that AC tend to make les spontaneous looking for social stimuli even when no false belief attribution is required. Hence, atypical ToM development in AC may be based on atypical pattern of social orienting, which is hypothesized to be caused by atypical functioning of subcortical structures such as amygdala and its communication with cortical structures [68, 69, 70]. Moreover, between 15 and 55% of experimental groups of children with autism do pass the elicited-response false-belief task but there is no evidence of a propensity to engage with other people in an *other-person-centered* way in children with autism. Rather they show marked abnormalities in social relatedness features such as: (i) limitations in relational and communicational role-taking; (ii) ill-organized feelings towards and for other; (iii) restrictions in self-awareness. Indeed, gestures, body language and facial expressions are poorly coordinated [71] and comprehension of nonverbal cues accompanying communication such as intonation, appear to be impaired [72]. In the Sticker Test [73], children without autism would often employ a point-to *themselves* to communicate that a tester should place a sticker on *herself*. They appeared to identify with the tester, and presume she in turn would identify with them pointing-to-themselves and place the sticker on her own (i.e., the tester's) body. Participants with autism seldom adjusted their communication in this mutually coordinated, person-anchored way. Instead most pointed directly to the tester's body. The RSI hypothesis is also in line with the observation that high-functioning individuals with autism having written auto-biographies [74, 75] may have much greater difficulties of expression in online conversational settings.

The lack of affective engagement is even more striking and well documented: for instance, Hobson et al. [7] interviewed parents with AC aged between 6-13 years. Although they could recognize in their children emotions such as anger, fear, etc., they rarely cited clear instances of other-person-centered emotions such as guilt, shame, pity or embarrassment. The RSI is also compatible with findings revealing that AC display limited propensity to register and assimilate the bodily-expressed attitudes of others in such a way as to apprehend and relate to the world through other people's minds and emotions. Also they exhibit limitations in perceiving and/or adjusting to the perspectives of different figures within a story, and in shifting from one-person anchored perspective to another [61].

II) At the *neural* level, fMRI studies show that one area in the brain crucial for making a distinction between self and other (i.e. the ventromedial prefrontal cortex vMPFC) responds atypically when individuals with autism reflect on themselves or others, namely it responds in an egocentrically equivalent way for both self and other [76, 53]. This lack of neural self-other distinction supports behavioral evidence [77, 78, 79] illustrating that that AC show markers of atypical self-other distinctions. Similarly, Chiu and colleagues [80] found that an area outside the traditional mentalizing system (the middle cingulate cortex; MCC) was specifically underactive when participants with ASC had to decide how much to invest in the other person while playing the 'trust game'[7] with another person.

6 CONCLUSIONS & FUTURE WORK

I argued here that the intersubjective relatedness of the self is foundational for and not merely subsidiary to the development of representational skills. That is, the lack of metarepresentational abilities may be the *effect* and *not* the cause of interpersonal relatedness impairments. Recent studies suggested that individuals with autism may have greater difficulty representing their own beliefs than the beliefs of other people [83]. Hence, an embryonic and speculative working hypothesis is that the ability to socially interact and read others' minds may have evolved *first*, with the turning inward of the metarepresentational spotlight upon our own inner states developing only later [84]. If this line of reasoning is correct, then it has profound implications for our understanding of minds not as mechanisms situated in individuals' heads, but rather as *constitutively* dependent upon worldly and social interactions. Indeed, Mindblindness theorists, in their search to establish the unshakeable *basis* of cognition and its explanatory *building blocks*, may have overestimated the role of conceptual thinking and underestimated the role of intrinsic social-emotional organization. Rather, the epistemic situation of a typically developing infant is, to use O. Neurath' metaphor, similar to that of a sailor who has to rebuild its ship incessantly, on the open sea, without ever being able to dismount it in dry-dock and reconstruct it from the best components, as theorists usually do. Hence, a crucial feature they benefit from is the constitutive and interactive *autonomy* that living systems enjoy by virtue of their self-generated identity as distinct entities in constant material flux. By contrast, failures in the ability to direct expressions of affect to another person, use facial expressions communicatively, and resonate to the emotions and bodily expressions of others suggest that autistic individuals act as *isolated* selves building a an isolated ship.

REFERENCES

[1] Baron-Cohen, S., Leslie, A. M., & Frith, U. (1985). Does the autistic child have a ''theory of mind''? *Cognition*, 21, 37–46.

[2] Leslie, A. M. (1994). "ToMM, ToBy, and Agency: Core architecture and domain specificity". In L. A. Hirschfeld & S. Gelman (Eds.), *Mapping the mind: Domain specificity in cognition and culture* (pp. 119–148). New York: Cambridge University Press.

[3] Hobson, R Peter (2012) "Emotion as personal relatedness." *Emotion Review*, 4 (2). pp. 169-175

[4] Schlick, M. (1934/1979) 'The Foundation of Knowledge', in M. Schlick, *Philosophical Papers Vol. II (1925–1936)*, H. L. Mulder and B. van de Velde-Schlick (eds.), Dordrecht: Reidel

[5] Kanner, L. (1943). Autistic disturbances of affective contact. *Nervous Child*, *2*, 217–250

[6] Chiu, P. H., Kayali, M. A., Kishida, K. T., Tomlin, D., Klinger, L. G., Klinger, M. R., et al (2008). "Self responses along cingulate

[7] The 'trust game' is a social interaction where emergent properties of the interaction play a major role in shaping the neural mechanisms elicited in both the investor and trustee's brain [81] but do not shape the same neural mechanisms when the interaction is non-social, such as when playing against a computer [82].

cortex reveal quantitative neural phenotype for high-functioning autism". *Neuron*, 57, 463–473.
[7] Hobson, R. P., Chidambi, G., Lee, A., & Meyer, J. (2006). Foundations for self-awareness: An exploration through autism. *Monographs of the Society for Research in Child Development*, *284*, 1–165.
[8] Charman, T., Swettenham, J., Baron-Cohen, S., Cox, A., Baird, G., & Drew, A. (1997). "Infants with autism: An investigation of empathy, pretend play, joint attention, and imitation. *Developmental Psychology*, *33*, 781–789.
[9] Sigman, M. D., Kasari, C., Kwon, J. H., & Yirmiya, N. (1992). "Responses to the negative emotions of others by autistic, mentally retarded, and normal children". *Child Development*, *63*, 796–807.
[10] Hobson, R. P. The emotional origins of social understanding. *Philosophical Psychology*, *6*, 227–249. (1993).
[11] Hobson, R. P. (2002). *The cradle of thought*. London, UK: Macmillan.
[12] Wimpory, D. C., Hobson, R. P., Williams, J. M., & Nash, S. "Are infants with autism socially engaged? A study of recent retrospective parental reports. *Journal of Autism and Developmental Disorders*, *30*, 525–536. (2000).
[13] Frith, U. Autism and "Theory of Mind". In C. Gillberg (Ed.), Diagnosis and Treatment of Autism. (pp. 33-52). New York: Plenum Press (1989).
[14] Leekam, S., Baron-Cohen, S., Perrett, D., Milders, M., & Brown, S. "Eye-direction detection: A dissociation between geometric and joint attention skills in autism". *British Journal of Developmental Psychology*, 75, 77-95. (1977).
[15] Leekam, S., Baron-Cohen, S., Perrett, D., Milders, M., & Brown, S. (1977). "Eye-direction detection: A dissociation between geometric and joint attention skills in autism". *British Journal of Developmental Psychology*, 75, 77-95.
[16] Ozonoff, S., Pennington, B. F., & Rogers, S. J. (1991). "Executive function deficits in high-functioning autistic individuals: Relationship to theory of mind". *Journal of Child Psychology and Psychiatry and Allied Disciplines*, *32*, 1081-1105.
[17] Meltzoff, A. N., & Gopnik, A. (1993). "The role of imitation in understanding persons and developing a theory of mind". In S. Baron-Cohen, H. Tager-Flusberg, & D.J. Cohen (Eds.), *Understanding other minds: Perspectives from autism* (pp. 335-365). Oxford: Oxford University Press.
[18] Gergely, G., & Watson, J. (1999). "Early social-emotional development: Contingency perception and the social biofeedback model. In P. Rochat (Ed.), *Early social cognition: Understanding others in the first months of life* (pp. 101–137). Hillsdale, NJ: Erlbaum.
[19] Gopnik, A., & Wellman, H. M. (1994). The theory theory. In L. Hirschfield, & S. Gelman (Eds), Mapping the mind: Domain specificity in cognition and culture (pp. 257-293. xiv, 516). New York: Cambridge University Press.
[20] Leslie, A. 2000: 'Theory of mind' as a mechanism of selective attention. In M. Gazzaniga (ed.), *The New Cognitive Neurosciences*, 2nd edn. Cambridge, MA: MIT Press.
[21] Perner, J., and T. Ruffman. 2005. "Infants' insight into the mind: how deep?" *Science* 308: 214–216.
[22] Goldman, A.I. (2006). *Simulating minds: The philosophy, psychology, and neuroscience of mindreading*. New York: Oxford University Press.
[23] Bermúdez, J. L. (2003). The domain of folk psychology. In A. O'Hear (Ed.), Minds and persons (pp. 25– 48). Cambridge: Cambridge University Press
[24] Gallagher, S. (2005). How the body shapes the mind. USA: Oxford University Press.
[25] Gallagher, S. (2009). Neural simulation and social cognition. In J. A. Pineda (Ed.), Mirror neuron systems (pp. 1–17). New York: Humana.
[26] Hutto, D. D. (2004). The limits of spectatorial folk psychology. Mind & Language, 19(5), 548–573.
[27] Hutto, D. D. (2008). Folk psychological narratives: the sociocultural basis of understanding reasons. Cambridge: MIT Press.
[28] Lakoff, G., & Johnson, M. (1999). Philosophy in the flesh: the embodied mind and its challenge to western thought. New York: Basic Books.
[29] De Jaegher, H. & E. Di Paolo. (2007). 'Participatory sense-making: An enactive approach to social cognition'. *Phenomenology and the Cognitive Sciences* 6: 485–507
[30] Godfrey-Smith, P. (2005). Folk psychology as a model. *Philosophers' Imprint*, *5*(6).
[31] Spaulding, S. (2010), 'Embodied Cognition and Mindreading'. *Mind & Language*, 25: 119–140.
[32] Gallagher, S., (2011) 'In Defense of Phenomenological Approaches to Social Cognition: Interacting with the Critics', *Review of Philosophy and Psychology*, June 2012, Volume 3, pp 187-212
[33] Meltzoff, A. N., & Moore, M. K. (1977). Imitation of facial and manual gestures by human neonates. *Science*, *198*, 75-78.
[34] Meltzoff, A. N., & Moore, M. K. (1989). Imitation in newborn infants: Exploring the range of gestures imitated and the underlying mechanisms. *Developmental Psychology*, *25*, 954-962.
[35] Ekman, P., Friesen, W. V., & Ellsworth, P. (1972). *Emotion in the human face*. New York: Pergamon Press.
[36] Ekman, P., Levenson, R. W., & Friesen, W. V. (1983). "Autonomic nervous system activity distinguishes between emotions". *Science, 221*, 1208-1210.
[37] Izard, C. E. (1977). *Human Emotions.* New York: Plenum Press.
[38] Izard, C. E. (1978). Emotions as motivations: An evolutionary-developmental perspective. In H. E. Howe, Jr. (Ed.), *Nebraska Symposium on Motivation* (Vol. 26, pp. 163-199). Lincoln: University of Nebraska Press.
[39] Bruner, J. S., Olver, P. R., & Greenfield, P. M. (1966). *Studies on cognitive growth.* New York: Wiley.
[40] Birch, H. G., & Lefford, A. (1967). "Visual differentiation, intersensory integration, and voluntary control". *Monographs of the Society for Research in Child Development, 32* (Serial No. 110).
[41] Gholson, B. (1980). *The Cognitive-Developmental Basis of Human Learning: Studies in Hypothesis Testing.* New York: Academic Press.
[43] Goldman, A.I. (2005). "Imitation, mind reading, and simulation. In Perspectives on imitation". In II, ed. S. Hurley and N. Chater, 80–91. Cambridge: MIT Press.
[44] Rizzolati, G., & Craighero, L. (2004). The mirror-neuron system. *Annual Review of Neuroscience, 27*, 169–192.
[45] Gallese, V., & Goldman, A., (1998), "Mirror neurons and the simulation theory of mind-reading," *Trends in Cognitive Sciences*, 2: 493–501.
[46] Gallese, V. (2010) 'Embodied Simulation and its Role in Intersubjectivity' in Fuchs, Th., Sattel, H., Henningsen, P., (eds), *The Embodied Self*, Stuttgart, Schattauer.
[47] Williams, J. H., Whiten, A., Suddendorf, T. & Perrett, D. I. Imitation, mirror neurons and autism. *Neurosci. Biobehav. Rev.* 25, 287–295 (2001).
[48] Iacoboni, M., & Dapretto, M. (2006). The mirror neuron system and the consequences of its dysfunction *Nature Reviews Neuroscience*, 7, 942–951.
[49] Catmur, C., Walsh, V. & Heyes, C. (2007). Sensorimotor learning configures the mirror neuron system. *Current Biology*, 17(**17**): 1527-1531.
[50] Hutto, D. & Myin, E. (2013), *Radicalizing Enactivism*, Cambridge, MA : MIT Press.
[51] Neisser, U., 1988. "Five kinds of self-knowledge," *Philosophical Psychology*, 1: 35–59.
[52] Rochat, P., & Striano, T., (2000). "Perceived Self in Infancy", Infant Behavior and Development, Vol. 23, Issue 3-4, pp. 513–530.
[53] Lombardo, M. V., & Baron-Cohen, S. (2010), "The role of the self in mindblindness in autism". *Consciousness and Cognition* (2010), doi:10.1016/j.concog.2010.09.006
[54] Hobson, R. P., García-Pérez, R., & Lee, A. (2009). Person-centred (deictic) expressions and autism. *Journal of Autism and Developmental Disorders*, *40*, 403–415.

[55] Aitken, K.J., & Trevarthen, C. (1997). Self–other organization in human psychological development. Development and Psychopathology, 9, 651–675.
[56] Tzourio-Mazoyer, N., De Schonen, S., Crivello, F., Reutter, B., Aujard, Y., & Mazoyer, B. (2002). Neural correlates of woman face processing by 2-month-old infants. NeuroImage, 15, 454–461.
[57] Jaffe J, Beebe B, Feldstein S, Crown CL, Jasnow MD. (2001), Rhythms of dialogue in infancy: coordinated timing in development. *Monographs of the Society for Research in Child Development*. 2001;66 (2):i-viii, 1-132.
[58] Trevarthen, C. (2005). Stepping away from the mirror: Pride and shame in adventures of companionship – reflections on the nature and emotional needs of infant intersubjectivity. In C.S. Carter, L. Ahnert, K.E.H. Grossman, S.B.M.E. Lamb, S.W. Porges, et al. (eds.), Attachment and bonding: A new synthesis (Dahlem Workshop Report 92). (pp. 55–84). Cambridge, MA: MIT Press.
[59] Hermelin, B., & O'Connor N. (1967), "Remembering of Words by psychotic and subnormal children". *British Journal of Psychology*, 58, 213-218.
[60] Tager-Flusberg, H. (1992). "Autistic children's talk about psychological states: deficits in the early acquisition of a theory of mind". *Child Development*, 63, 161–172.
[61] *Garcia-Pérez, R.*, Hobson, R. P., Lee, A., (2008), " *Narrative* Role-*Taking* in Autism*"* in *Journal of Autism and Developmental Disorders*, 38(1):156-68.
[62] Klin, A., Volkmar, F. R., & Sparrow, S. S. (1992). "Autistic social dysfunction: Some limitations of the theory of mind hypothesis". *Journal of Child Psychology and Psychiatry*, 33, 861–876.
[63] Wimmer, H. and J. Perner, 1983, "Beliefs About Beliefs: Representation and Constraining Function of Wrong Beliefs in Young Children's Understanding of Deception," *Cognition*, 13: 103–128.
[64] Onishi, K. and Baillargeon, R. 2005: Do 15-month-old infants understand false beliefs? *Science*, 308, 255–8.
[65] Senju, A., Southgate, V., Miura, Y., Matsui, T., Hasegawa, T., Tojo, Y., et al (2010). "Absence of spontaneous action anticipation by false belief attribution in children with autism spectrum disorder". *Development and Psychopathology*, 22, 353–360.
[66] Turner, M. A. (*1996*). *Repetitive behaviour* and cognitive functioning in autism. Unpublished PhD thesis, University of Cambridge, Cambridge, UK
[67] Ruffman, T., Garnham,W.,& Rideout, P. (2001). "Social understanding in autism: Eye gaze as a measure of core insights". *Journal of Child Psychology and Psychiatry*, 42, 1083–1094.
[68] Dawson,G., Webb, S. J.,&McPartland, J. (2005). "Understanding the nature of face processing impairment in autism: Insights from behavioral and electrophysiological studies". *Developmental Neuropsychology*, 27, 403–424.
[69] Johnson, M. H. (2005). "Subcortical face processing*". Nature Reviews Neuroscience*, 6, 766–774.
[70] Schultz, R. T. (2005). "Developmental deficits in social perception in autism: The role of the amygdala and fusiform face area". *International Journal of Developmental Neuroscience*, 23, 125–141.
[71] Tantam, D. (1988b) Lifelong eccentricity and social isolation II: Asperger's syndrome or schizoid personality disorder? British Journal of Psychiatry, 153, 783-791.
[72] Hobson, R.P. (1986). "The autistic child's appraisal of expressions of emotions", J Child Psychol Psychiatry, 27; 321-42.
[73] Hobson, R. P., & Meyer, J. A. (2005). "Foundations for self and other: A study in autism". *Developmental Science*, *8*, 481–491.
[74] Grandin, T. (1984), "My experiences as an autistic child and review of selected literature". J Orthomolecular Psychiatry, 13: 144-75.
[75] Williams D. (1992), *Nobody nowhere*. Toronto: Doubleday Canada.
[76] Kennedy, D. P., & Courchesne, E. (2008). Functional abnormalities of the default network during self- and other-reflection in autism. Social Cognitive and Affective Neuroscience, 3, 177–190.
[77] Lee, A., & Hobson, P. R. (2006). "Drawing self and others: How do children with autism differ from those with learning difficulties". *British Journal of Developmental Psychology*, 24, 547–565.
[78] Lee, A., Hobson, R. P., & Chiat, S. (1994). "I, you, me, and autism: An experimental study". *Journal of Autism and Developmental Disorders*, 24, 155–176.
[79] Mitchell, P., & O'Keefe, K. (2008). "Brief report: Do individuals with autism spectrum disorder think they know their own minds?" *Journal of Autism and Developmental Disorders*, 38, 1591–1597.
[80] Chiu, P. H., Kayali, M. A., Kishida, K. T., Tomlin, D., Klinger, L. G., Klinger, M. R., et al (2008). "Self responses along cingulate cortex reveal quantitative neural phenotype for high-functioning autism". *Neuron*, 57, 463–473.
[81] King-Casas, B., Tomlin, D., Anen, C., Camerer, C. F., Quartz, S. R., & Montague, P. R. (2006). "Getting to know you: Reputation and trust in a two-person economic exchange". *Science*, 308, 78–83.
[82] Tomlin, D., Kayali, M. A., King-Casas, B., Anen, C., Camerer, C. F., Quartz, S. R., et al (2006). "Agent-specific responses in the cingulate cortex during economic exchanges". *Science*, 312, 1047–1050.
[83] Williams D.M., Happé. F (2009), 'What did I say? Versus what did I think? Attributing false beliefs to self amongst children with and without autism'. *Journal of Autism and Developmental Disorders* June 2009, Volume 39, Issue 6 pp 865-873.
[84] Happé, F. (2003). Cognition in autism: one deficit or many? *Novartis Foundation Symposium, 251*, 198-207.

Autism as Philosophical Insight: The Enactive Response to the Tendency to Pathologize

Joel Parthemore [1]

Abstract. An increasing trend in recent years has seen the pathologizing of a wide range of mental phenomena under the headings of "disease", "disorder", “illness”, or "disability". A recent school in philosophy of mind, enactivism, is inclined to oppose this trend as based upon philosophically dubious assumptions. In particular, enactivism rejects mind-brain reductionism and simple emergentist accounts. It emphasizes the continuity between agent and environment, co-created out of the interaction. Enactivism is inclined to view much if not most of autism spectrum "disorders" as instances of cognitive diversity, recasting the disabilities associated with autism as a problem, not of neurology, but of the relationship between autist and society. The result is a failure to address the full range of human needs and talents and the mislabeling of the presumed minority as "broken". This paper shows how a re-conceptualization of autism can be not only philosophically better justified but also beneficial to autists and public alike.

1 INTRODUCTION

The diagnosis of autism and autism spectrum disorders dates to the 1940s. Until relatively recently, they were considered extremely rare; but, as Newschaffer and colleagues [14] write, “today, the prevalence of ASDs is understood to be many times greater, with the condition now thought to be second only to mental retardation among the most common developmental disabilities in the United States”. The growth in autism diagnoses is far from unique, if one considers other conditions falling under the broad heading of mental health. In 1988, there were 500,000 reported cases of attention deficit hyperactivity disorder in the States. By 2007, that number had risen to 4,000,000.

Various reasons have been put forward for these trends: not least the possibility that these diseases or disorders are more prevalent than previously thought; or that they are, for less than fully understood reasons, becoming more prevalent. At the same time, various people have noted the financial incentive for the pharmaceutical companies to encourage the labeling of phenomena as diseases requiring medication, with all the more profit to be made from chronic diseases requiring long-term medication. Others would put the finger more broadly on the politics of healthcare funding: a definite diagnosis with a clear treatment plan comes with a much more predictable cost.

What is most interesting for my purposes, however, are the often unstated philosophical – often, specifically metaphysical – assumptions that lie behind contemporary professional and public understanding of mental health issues. Modern science has taken a principled stand against supernatural explanations and against Cartesian divisions of mental thought from physical matter. All phenomena, it is said, have a natural and physical explanation: positions known respectively as naturalism and physicalism. Such a position might or might not be valid. What is significant for my purposes is the particular way in which naturalism and physicalism are frequently if not generally cashed out – not assisted by the common failure to define either term but rather take them as understood.

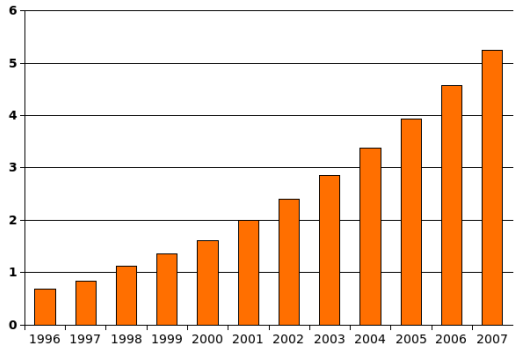

Figure 1. Bar chart of the number (per 1,000 U.S. resident children aged 6–17) of children aged 6–17 who were served under the Individuals with Disabilities Education Act (IDEA) with a diagnosis of autism, from 1996 through 2007. Reproduced under Creative Commons License from http://commons.wikimedia.org/wiki/File:US-autism-6-17-1996-2007.png.

This raises two problems. Mental phenomena get reductively treated as just another form of physical phenomena, setting up what I claim is a false equivalence between physical disorders and mental ones. The very term “mental illness” implies this. Illnesses are physically localizable phenomena, whereas I argue that mental phenomena simply are not, in general, localizable in the same way as physical ones. Illnesses come with a certain script: when one is ill, one goes to the doctor for treatment; one receives drugs or is given some specific treatment plan; one hopes for a cure or at least an amelioration of symptoms. If mental health issues are, indeed, a different order of things, then they call for a different set of scripts.

The muddled understanding of mind’s relationship to matter gives rise to further conceptual confusions: in particular, a frequent tendency to conflate things that legitimately are disorders from things that are not. Autism may not be considered an illness, but it is commonly referred to as a disorder and treated as a disability (an even more vague and problematic term). All of these terms – disorder, illness, disability – are loaded with conceptual baggage. I argue that much of what is considered to fall within autism spectrum disorders – as well as other phenomena that are not considered diseases but are

[1] Centre for Cognitive Semiotics, University of Lund, Sweden; email: joel.parthemore@semiotik.lu.se.

considered disorders – should better be understood in more neutral terms, as expressions of cognitive diversity, where that diversity is *ceteris paribus* a potential benefit and not liability to society.

Several caveats are in order. First, I am neither a psychologist nor psychiatrist, and nothing I say here should be taken to imply otherwise. I am a philosopher of mind with a particular interest in how people organize their conceptual understanding of the world, especially their ability to conceptualize the "same things" in multiple, equally valid but often mutually exclusive, ways. I ascribe to a school of philosophy known as enactivism that implies a different, albeit decidedly non-Cartesian, understanding of the relationship of cognitive and biological self: one that has been well explored in some domains but not when it comes to issues of mental health and cognitive dysfunction. Second, I am not claiming that autism does not exist nor that it is, as some radical constructivists might have it, a purely social construction. Finally, I am certainly not denying that autism, as with all mental phenomena, has a substantive physical basis, nor that there are not forms of autism that are legitimately understood as disorders. Rather, my concern is with how autism (broadly understood) is conceptualized – both professionally and publicly – as well as with how that conceptual structure then prejudices how people who are labeled with autism spectrum disorders get treated.

My primary interest in writing this paper is to demonstrate the practical consequences of adopting an enactive perspective. One of the frequently heard complaints about Andy Clark and Dave Chalmers' extended-mind hypothesis [5, 4] – in some ways, a close cousin of enactivism – is the lack of practical benefit to adopting the position: see e.g. [25, p. 15]. People understandably want to see how empirical research proceeds any differently. If it does not, then the philosophy looks like so much armchair theorizing. I think that mental health is a particularly fruitful area for applying enactive philosophy; this paper will attempt to explain why. I focus my attention on autism because, although some mental health issues are untendentiously labeled disorders, the labeling of autism spectrum phenomena as disorders has, in many cases, been hotly debated.

I proceed from the conviction – argued for in e.g. [18, 16, 17, 20] – that, as a general rule, conceptual frameworks do not have privileged access to the truth in any ontologically prior sense. Change the conceptual framework and, in a very real way, you change the world. Change the way mental health issues are conceptualized, and you will unavoidably change how they are addressed.

I proceed as well from personal motivations, having had close friends who were diagnosed Asperger's (a separate diagnosis from autism spectrum disorders in DSM-IV[2] but included under autism in DSM-V). Although I have never been tested in a clinical setting, I myself score quite high on one of the standard tests used to diagnose Asperger's. I have close friends and family who have been diagnosed at one time or another with various other illnesses, diseases, or disorders, including schizophrenia, manic-depressive disorder, and depression. If any or all of these can or even should be approached differently, I want to understand how.

Section Two introduces *enactivism*, a school of philosophy with roots in the work of Humberto Maturana and Francisco Varela from the 1970s onward, with its distinctive, co-creational view of the relationship between agent and environment. Section Three sets forth the argument for an alternative conceptualization of mental "disorders" in general and autism spectrum disorders in particular, contrasting it with prevailing views in American and European psychiatry. Section Four considers the consequences of such a radical reconceptualization and offers some prescriptions for the way forward.

Along the way, there are a number of pitfalls I would seek to avoid. I suggest labeling many disorders – including high-functioning autism and Asperger's – as instances of cognitive diversity. Yet, for all my desires to avoid conceptual baggage, "diversity" is, itself, a loaded word, scorned by many for its supposed political correctness. I aim to attack certain – what I see as naive – *naturalist* and *physicalist* positions. At the same time, I want to avoid any endorsement of non-physicalism, supernaturalism, or metaphysical idealism.

Some would call autism the easy case. A too-cursory reading of this paper might suggest the conclusion that mental disorders do not exist: a position that certainly has been argued for, most notably by Thomas Szasz [27], but is not one I wish to adopt here. Finally, others would say that my position is just social psychiatry recycled. I need to show what is new here, beyond just the *language* of enactive philosophy.

2 ENACTIVISM

Associated in its origins with people such as Francisco Varela, Humberto Maturana, Evan Thompson, and Eleanor Rosch and with books such as *The Embodied Mind* [29] and *The Tree of Knowledge [13],* enactivism emphasizes the interaction of the individual with her environment in all sorts of boundary-transgressing ways; and much is made – if not always well explained – about the co-creation of agent and environment. As the poet Antonio Machado writes, "wanderer, the road is your footsteps, nothing else; wanderer, thee is no path, you lay down a path in walking" (quoted in [28, p. 63]).

It endorses notions of *situatedness* (an agent is always embedded in a particular physical and – for some agents – social environment), *embodiment* (an agent always takes a particular physical form), and *extended mind* (mind extends into the world in a way that body does not), even while it tends to complain that none of these, on their own, go far enough.[3] It reserves particular criticism for extended mind for suggesting that cognition is mostly in the head when, on an enactive account, cognition is not *in* the head at all but "in" the interaction: i.e., cognition is physically realized, but it is not, itself, a physical entity (with e.g. dimensions of length, width, height, volume, mass). Neither can it be reduced to one nor can any straightforward story be told of how arises from one.

Rather – on reflection – it is the product of a different perspective to the usual perspective of empirical science, in which the role of the observer downplayed, ignored, or denied altogether. When it comes to matters of mind, consciousness, cognition – wherever one looks, the observer is there, from the beginning. As Maturana writes [12, p. 30], "*everything that is said, is said by an observer to another obserer that could be himself.*" The irreducible role and influence of the observer is key: the observer changes the observed merely by her presence.

As I argue in [19], it is part and parcel of human cognition to toggle incessantly between these two perspectives on the world: one that we might call "mental", the other "physical". Both are necessary even as they are ultimately unreconcilable into a single, joint perspective. Thus, only one can be the focus of our attention at any given time.

With its tendency toward practical if not necessarily metaphysical antirealism – by which I mean the inseparability, in practice or in principle, of mind from mind-independent world, for any agent who has a mind – enactive philosophy poses a number of challenging questions to contemporary popular and clinical treatment of mental

[2] The American Psychiatric Association's Diagnostic and Statistical Manual of Mental Disorders.

[3] Contrast this with e.g. Robert Rupert [25, 26, 24], for whom embeddedness and embodiment is precisely where one should *stop.*

health issues in general and – the focus of this paper – autism spectrum disorders in particular.

Is autism legitimately labeled a disorder? If it is *sometimes* a disorder, is it *always* one? What does it mean to be a disorder, anyway? Where is the disorder located? (Whose disorder is it?) If it is to be viewed – at least sometimes – as e.g. cognitive diversity rather than disorder, how does that change how we treat it? What is the appropriate role of neurology – for enactive philosophy certainly agrees that it has one – and to what extent can it meaningfully be separated from the rest of the holistic entity that is an individual, when it comes to matters of mental health? That is, do *mental* phenomena, unlike straightforwardly *physical* phenomena, press firmly toward a holistic rather than reductionist account?

3 RECONCEPTUALIZING AUTISM: THE ARGUMENT FROM ENACTIVISM

To the enactive perspective, nearly everything comes back to interaction. Life is movement, and meaning arises out of interaction: the interaction of an agent with her physical and social environment. Interaction is a dynamic, ever-changing process that varies from individual to individual. This has two consequences for conceptual frameworks:

1. If movement is life, then stasis is death. No conceptual framework can be immune to change. To the extent that a given framework *stops* changing, it stops being applicable to a world that *is* changing. To the extent it denies the change it is undergoing, it becomes opaque to itself.
2. If no two interactions are ever precisely the same – if interaction is not, ultimately, the sort of thing that can be reified anyway – then, at least for most domains, no conceptual framework can ever claim to be the *right* framework, and one must allow the possibility of multiple – possibly mutually exclusive – frameworks each with some degree of validity. The question becomes: what are the practical advantages and disadvantages of going with one framework rather than another?

Autism spectrum disorders are all about an autist's supposed lack of interaction – or lack of "normal" interaction – with her environment. For the same reason there can be no "brain in a vat"interacting only with itself, entirely cut off from its environment – here I refer the reader to Hilary Putnam's classic thought experiment [22] – so, too, enactivism renders the notion of an agent removed from (interaction with) an environment incoherent. The autist's problem – to the extent that there *is* a problem – cannot be lack of interaction but the type of interaction. Depending on where one falls on the autism spectrum – depending on whether one has the now deprecated diagnosis of Asperger's instead – typically cited signs of autism in children include[4]:

- Difficulty mixing with other children.
- Difficulty expresing needs, often using gestures instead of words.
- Inappropriate laughing or giggling.
- Inappropriate attachment to objects.
- Lack of eye contact.
- Lack of fear of danger.
- Apparent insensitivity to pain.
- Apparent pleasure in spinning objects.
- Preference for being alone.
- Over- or underactivity.
- Unresponsiveness to normal teaching methods.
- Unresponsiveness to verbal cues; tendency to act deaf.
- Insistence on routines.
- Sustained odd play.
- Avoidance of physical contact, such as cuddling or hugging.
- Uneven motor skills.

Figure 2. Quinn, a boy of approximately 18 months, seen intently stacking cans. Reproduced under the Creative Commons License from http://upload.wikimedia.org/wikipedia/commons/c/ca/Autism-stacking-cans_edit.jpg.

What matters is not the precise contents of the list but the overall picture they create: social awkwardness crossing over, in extreme cases, to dysfunction; lack of attention to social cues; difficulty with socially complex environments; perception to others of "not fitting in"; and so on. The autist's *social* environment is key; and it is to the social environment one must look for any better understanding of autism.

To answer the question, is autism legitimately labeled a disorder?, one must first, of course, decide what *is* a disorder. With its emphasis on engagement and interaction, enactivism suggests a practical approach: is the autist – at least, the adult autist – able to meet her basic physical and emotional needs or not? What is her self-measure of life satisfaction? If she is happy with her life and she is able to meet her basic needs then, enactivism suggests, she does not have a disorder.

Misconceptions about autism abound – and not only in the general public. There is a widespread conception that autists suffer greatly and are generally unhappy people. Although this may be true as a general rule – and longitudinal research seems to support this – there clearly is a great deal of individual variation, even in cases of low-functioning autism; as Audrey Burgess and Steven Gutstein acknowledge [3, p. 80], "QoL [quality-of-life] research for people with autism is lacking." As both they and Jo Renty and Herbert Roeyers [23] report, much has to do with how the social environment engages with the autist. As Renty and Roeyers write [23, p. 511], in somewhat dry, technical language, "support characteristics are related to quality of life in adults with ASD, whereas disability characteristics are not." If autists are unhappy people, they are not *intrinsically* so.

[4] I an uncertain of the original source on this list, nearly matching versions of which can be found in many places in the literature and on the Internet, e.g. http://www.asaetc.org/asahome/?page_id=416.

At the same time, it is commonly believed that autists do not need other people and thus do not experience loneliness; clearly, this is not the case, either: so e.g. Nirit Bauminger and Connie Kasari [2] report that (high-functioning) autists show *higher* levels of loneliness than the background population. What gets taken as a lack of interest in social engagement may simply be a failure to know *how* to communicate: a failure that can, needless to say, go both directions.

It seems that part, if not much, of the time – given the appropriately supportive environment – autism spectrum phenomena need *not* be understood as disorders at all. Where they *are* appropriately labeled as disorders, an enactive perspective is inclined to locate the disorder neither in the agent nor in the environment but "in" the interaction – with responsibility, if you will, to be shared. Here is where an enactive perspective on mental health distinguishes itself from social psychiatry, broadly understood: if it sees the currently dominant biopsychiatric model as too focused on the agent to the exclusion of her environment: i.e., too internalist; then it sees social psychiatry as too focused on the environment, to the exclusion of the agent: i.e., too externalist.[5] Internalism and externalism are both extremes to be avoided, even if one is necessarily pulled, at any given moment, to one or the other perspective. Which perspective is currently appropriate depends very much on the questions one is asking and the applications to which one intends to put the answers. If one is looking for causal factors then – enactivism suggests – a holistic, environment-embracing perspective is needed, and the appropriate model of causality is likely circular rather than linear (*cf.* [20, p. 297], [16, pp. 142-143]). If one is looking for interactions with physical health – and one should always expect to find such interactions! – then a more reductionist perspective that downplays or even largely excludes the environment is more appropriate. Enactivism would in no way deny that there are significant differences in e.g. the brains of autists versus the background population; the question is rather how much these are causal and how much the product of the autist's peculiar form of interaction.

The enactive criticism of the biopsychiatric model – as I am portraying it – rests on three key points: what I call the *problem with physicalism,* the *problem with naturalism*, and the *problem with dogmatism.*

3.1 The problem with physicalism.

Sure: *ceteris paribus,* substance monism is to be preferred over substance dualism. It is difficult to find anyone who clearly endorses a substance dualist account: it is controversial among Descartes scholars whether Descartes himself did (see e.g. [1, 15]).

One can call that substance "physical" if one likes; but how well do we grasp what it means to be physical stuff, except in a rough-and-ready everyday way? As with my reading of enactivism, the neutral monist – and here one finds such names cited as Spinoza, William James, and Bertrand Russell – is inclined to see *one* substance viewed from necessarily *two* perspectives: necessarily two, because of our inability to step outside of our observational system to "see things as they really are". In particular, an enactive perspective is inclined to question the common assumption that the relevant physically grounded explanations are necessarily physically localizable. If mind extends substantively into world, they very well may not be.

[5] Indeed, this is my criticism of Alva Noë's brand of enactivism in [20].

3.2 The problem with naturalism.

If possible, even less effort is made to define naturalism than physicalism. Assumptions that "everyone knows" what one means may be dangerous. In most uses, it seems to amount to something like "able to be given a complete and consistent account, with no explanatory residue". If Douglas Hofstadter (see e.g. [6, 7]) – among many others – is right, and human cognition is knowably bounded even as the universe is not, then not everything *can* be "fully" naturalized (most particularly, perhaps, ourselves). Some things will be more amenable to naturalistic explanations than others, and some will resist any such explanations quite robustly. I believe in tesseracts and hyperspheres, but I do not think that makes me a supernaturalist! An approach to mental health that lets go of the need *always* to naturalize may be more open to an approach that is as much art as science.

3.3 The problem with dogmatism.

The current biopsychiatric model clearly assumes the possibility of one correct conceptual framework. I have already called that into question. The assumption that there can *be* one correct model just is one definition of dogmatism. Such an assumption is valid *if and only if* one can truly know that one is right (and not just know, but know that one knows). Enactivism shares the anti-dogmatic inclinations of American pragmatism – I have in mind such people as William James [8] and Charles Sanders Pierce [21] – and, going further back in time, the ancient Greek school of Pyrrhonism, a very fine account of which is offered by Per Lind [11]. Given my perspective on enaction, I find particularly appealing the Pyrrhonian solution to dogmatism, which is to exploit cognitive conflicts and contradictions to tear down the existing conceptual framework and allow a new one. The biopsychiatric model with its strongly internalist bias is, I believe, ripe for such attack.

4 THE FRUITS OF RADICAL RECONCEPTUALIZATION

As a generalization of Thomas Kuhn's [10, 9] notion of *paradigm shifts,* by which he described the wholesale conceptual change of scientific frameworks within the scientific community, I prefer the term *radical reconceptualization*: the wholesale replacement of any conceptual framework by another [18], something that the Pyrrhonists saw as an ongoing, iterative process. Like conceptual change on any level, the discontinuities implied by radical reconceptualization assume underlying continuities. It is never possible to throw *everything* out and start over; the trick is knowing what to discard and what to keep. In the broader context, radical reconceptualization is a powerful tool allowing us to reconceptualize our world by reconceptualizing ourselves, and vice versa. In the context of mental health issues, the time is ripe to take a fresh look at illnesses, disorders, and disabilities.

Enactivism offers a fresh approach in part inspired by, but going beyond, social psychiatry, with its reaction against the overly agent-centric model: one that permits a more flexible, more negotiable boundary between mentally well / "normal" and mentally deviant / disordered / diverse. In the case of autism spectrum disorders, it holds out the promise of attaching less stigma to, and facilitating more positive contributions to society from, in particular – but by no means exclusively – high-functioning autists and those diagnosed with Asperger's. Just as society should take seriously the paraplegic who insists that he is not disabled, it can see the autist not as a broken

individual but as one with a different way of looking at and engaging with the world. As the well-worn saying goes, autism is "**A**lways **U**nique, **T**otally **I**nteresting, **S**ometimes **M**ysterious."

REFERENCES

[1] G. Baker and K. Morris, *Descartes' Dualism*, Routledge, 2002.

[2] N. Bauminger and C. Kasari, 'Loneliness and friendship in high-functioning children with autism', *Child Development*, **71**(2), 447–456, (March/April 2000).

[3] A. F. Burgess and S. E. Gutstein, 'Quality of life for people with autism: Raising the standard for evaluating successful outcomes', *Child and Adolescent Mental Health*, **12**(2), 80–86, (2007).

[4] A. Clark, *Supersizing the Mind: Embodiment, Action, and Cognitive Extension*, Oxford University Press, 2008. Also available as ebook through Oxford University Press (http://www.oxfordscholarship.com).

[5] A. Clark and D. Chalmers, 'The extended mind', *Analysis*, **58**(1), 7–19, (January 1998).

[6] D. Hofstadter, *Gödel, Escher, Bach: An Eternal Golden Braid*, Penguin, 2000. Twentieth anniversary edition.

[7] D. R. Hofstadter, *Gödel, Escher, Bach: An Eternal Golden Braid*, Basic Books, Inc., 1979. Original publication.

[8] W. James, *The Meaning of Truth*, chapter 8, 180–216, Dover, 2002.

[9] T. Kuhn, *The Structure of Scientific Revolutions*, University of Chicago Press, Chicago, 1970.

[10] T. Kuhn, 'The road since Structure', in *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association*, volume 2 of *1990*, pp. 3–13, (1990).

[11] P. Lind, 'Pyrrhonian aporiagogics and the astrocognitive challenge', in *The History and Philosophy of Astrobiology: Perspectives on the Human Mind and Extraterrestrial Life*, eds., D. Duner, J. Parthemore, G. Holmberg, and E. Persson, chapter 4, Cambridge Scholars, (2013). In press.

[12] H. Maturana, 'Cognition', in *Wahrnehmung und Kommunikation*, eds., P. M. Hejl, W. K. Köck, and G. Roth, 29–49, Peter Lang, Frankfurt, (1978). Available online at http://www.enolagaia.com/M78bCog.html, with the original page numbering retained.

[13] H. R. Maturana and F. J. Varela, *The Tree of Knowledge: The Biological Roots of Human Understanding*, Shambhala, London, 1992.

[14] C. J. Newschaffer, L. A. Croen, J. Daniels, E. Giarelli, J. K. Grether, S. E. Levy, D. S. Mandell, L. A. Miller, J. Pinto-Martin, J. Reaven, A. M. Reynolds, C. E. Rice, D. Schendel, and G. C. Windham, 'The epidemiology of autism spectrum disorders', *Annual Review of Public Health*, **28**, 235–258, (April 2007). DOI: 10.1146/annurev.publhealth.28.021406.144007.

[15] M. Oakeshott, *Rationalism in Politics and Other Essays*, Liberty Press, Indianapolis, Indiana, USA, 1991.

[16] J. Parthemore, *Concepts Enacted: Confronting the Obstacles and Paradoxes Inherent in Pursuing a Scientific Understanding of the Building Blocks of Human Thought*, Ph.D. dissertation, University of Sussex, Falmer, Brighton, UK, March 2011. Available from http://www.sussex.ac.uk/Users/jep25/papers/thesis/thesis-a4.pdf.

[17] J. Parthemore, 'Of boundaries and metaphysical starting points: Why the extended mind cannot be so lightly dismissed', *Teorema*, **30**(2), 79–94, (2011).

[18] J. Parthemore, *The "Final Frontier" as Metaphor for Mind: Opportunities to Radically Reconceptualize What it Means to be Human*, chapter 3, Cambridge Scholars, 2013. In press.

[19] J. Parthemore, 'The unified conceptual space theory: An enactive theory of concepts', *Adaptive Behavior*, (2013). In press.

[20] J. Parthemore and A. F. Morse, 'Representations reclaimed: Accounting for the co-emergence of concepts and experience', *Pragmatics & Cognition*, **18**(2), 273–312, (August 2010).

[21] C. S. Peirce, 'How to make our ideas clear', in *The Nature of Truth: Classic and Contemporary Perspectives*, ed., M. P. Lynch, chapter 8, 193–209, MIT Press, (2001).

[22] H. Putnam, *Reason, Truth, and History*, Cambridge University Press, Cambridge, UK, 1981.

[23] J. Renty and H. Roeyers, 'Quality of life in high-functioning audlts with autism spectrum disorder', *Autism*, **10**(5), 511–524, (September 2006).

[24] R. Rupert, 'Challenges to the hypothesis of extended cognition', *Journal of Philosophy*, **101**, 389–428, (2004).

[25] R. Rupert, *Cognitive Systems and the Extended Mind*, Oxford University Press, 2009.

[26] R. Rupert, 'Critical study of Andy Clark's Supersizing the Mind', *Journal of Mind and Behavior*, **30**, 313–330, (2009).

[27] M. Thomas S. Szasz, *The Myth of Mental Illness: Foundations of a Theory of Personal Conduct (Revised Edition)*, Harper Perennial, 1984.

[28] F. J. Varela, 'Laying down a path in walking', in *Gaia: A Way of Knowing. Political Implications of the New Biology*, ed., W. Thompson, 48–64, Lindisfarne Press, Hudson, NY, USA, (1987).

[29] F. J. Varela, E. Thompson, and E. Rosch, *The Embodied Mind: Cognitive Science and Human Experience*, MIT Press, 1991.

Epistemic Actions in Attachment Relationships and the Origin of the Socially Extended Mind

Dean Petters [1] and **Everett Waters** [2]

Abstract. Attachment Theory describes how humans possess a strong innate predisposition to emotionally attach to familiar people around them who provide physical or emotional security. When infants learn to trust in intimate and enduring relationships they will tend to use their carers to extend their minds. Such cognitive extension is likely to impact mental health by helping infants reach the upper limit of their cognitive performance. Conversely, a caregiver's failure to gain 'epistemic trust' may act as a contributory risk factor for multiple later psychopathologies

1 ATTACHMENT RELATIONSHIPS AND THE EXTENDED MIND HYPOTHESIS

The extended mind thesis in cognitive science suggests that entities in the external environment may interact with internal mental processes in such a way that these entities can be seen as extensions of the mind itself [4, 5]. Cognitive extension is usually cast in terms of extending the mind onto inorganic objects in the environment, such as a mathematician doing their 'working' on paper. However, extension onto external agents (such as caregivers, teachers, friends or work colleagues) is not ruled out in this approach. For cognitive extension to involve extension of mental states onto an agent rather than an inorganic object, the agent must be strongly trusted, relied upon and accessible as an information provider [5].

There are many adult occupations where intense social interaction and information transfer occurs between workers jointly undertaking complex information critical tasks. When trust is built up over time or through training - in surgical teams, between police partners, or in small military groups, then individuals may use other people to extend their own cognition. However, such adult work related examples of social extension of cognition are not the focus of this paper. Instead this paper will concentrate on considering how examples of social extended cognition may occur when caregivers provide external cognitive support to infants and children (whilst also considering attachment partnerships between caregivers and adolescents and between adult romantic partners).

The idea that infants, older children and even adult attachment partners all look to their carers as information sources about the broader world is a familiar one. For example, social referencing and joint attention are both well studied topics in developmental psychology [3]. From the perspective of the socially extended mind, infant social referencing and joint attention between infant and carer may be seen as physical actions that make the infant's mental computations faster, more reliable or less effortful by actually incorporating the carer's help within those cognitive operations. Socially extended cognition in infant-carer dyads are likely to be asymmetric - with caregivers providing the extra support and the infant's cognition being extended. So, if a carer (as part of the environment) is coupled to an infant's cognitive system in the right way, they become part of the infant's mind. This approach is therefore treating the infant cognition as occurring 'outside' and well as 'inside the head'.

How well can the affordances provided by a carer match the infant's cognitive requirements? How well can carers support higher performance cognition in infants? How closely coupled can infant computational needs and carer response be? The psychoanalyst Donald Winnicott remarked that: *"There is no such thing as a baby, only a baby and a mother"* [14, page 39]. Winnicott's view of a closely integrated mother-infant dyad is consonant with the idea of carers extending the minds of infants they care for. What this means is that if the infant's ongoing computational needs are met by sensitive and timely support from his or her carer then we might say that the carers cognitive support has become part of the infant's extended mind.

Clark shows how language extends minds:

> *"First, the simple act of labelling the world opens up a variety of new computational opportunities and supports the discovery of abstract patterns in nature. Second, encountering or recalling structured sentences supports the development of otherwise unattainable kinds of expertise. And third, linguistic structures contribute to some of the most important yet conceptually complex of all human capacities: our ability to reflect on our own thoughts and characters and our limited but genuine capacity to control and guide the shape and contents of our own thinking"* [4, page 44]

From our current perspective of exploring how attachment figures may extend infant minds we can see that they may be viewed as fulfilling the same three roles that Clark ascribes to language. That is, caregivers can also help label and conceptualise, structure, and facilitate self-reflection.

In developing a framework to explain the origin of the socially extended mind, this paper draws upon Attachment Theory as an existing well-developed theory which is focused upon long term intimate social and emotional interactions, including how trust is formed and can be lost in close relationships. Attachment Theory is also based upon an information processing foundation which facilitates theorizing in terms of internal and external cognition. Serving as mind extension is clearly part of what attachment theorists mean when they say that an attachment figure enables one to 'live a bigger life' than would otherwise be possible. After showing how the information processing foundation for Attachment Theory leaves it well placed to be augmented with inclusion of socially extended mind mechanisms, this paper will assess how an Attachment Theory extended in

[1] University of Birmingham, UK, email: d.d.petters@cs.bham.ac.uk
[2] SUNY, Stony Brook, USA.

this way is relevant to mental health issues.

2 INTRODUCTION TO ATTACHMENT THEORY

Prior to the second war, medical and social scientists understood very little about the nature of human relationships and the security they can bring to an individual. Then, about the time that Alan Turing was making brilliant contributions to his country's security and the field of cybernetics was being formed, the child psychiatrist John Bowlby began publishing seminal insights into the nature of the child's tie to its carers and the origins of interpersonal security. Bowlby's attachment theory grew to include insights from numerous disciplines including psychoanalysis, ethology, cybernetics, cognitive and developmental psychology, and artificial intelligence [2]. The theoretical framework provided by Bowlby has structured and focused the empirical observations of attachment interactions in diverse contexts. These contexts include interactions between infants and children and their caregivers; between romantic partners in adolescence and adulthood; in typical and pathological populations; and observing attachment relations from a biological and comparative psychology perspective. Together they provide a rich description of attachment related affect, behaviour and cognitions, and information about the contexts in which they occur - information essential to building and evaluating computational attachment models and simulations [11].

Bowlby's initial research focus was towards understanding particular normative attachment related social and emotional phenomena such as: the separation distress exhibited by children when they or their mothers were absent due to the infant's or mother's hospitalization; the effect of early maternal deprivation on later development; and grief and mourning in infancy [2]. One of Bowlby's early goals was to construct a scientifically respectable attachment motivation theory that could account for an infant behaviour's sensitivity to social context. To accomplish this, he first turned to ethology and developed a framework that described the attachment system as an instinctive behaviour system. However, in his more mature theoretical work, Bowlby drew increasingly on control systems theory and on AI based representations such as internal working models (IWMs) and hierarchical plans. This means that the information processing explanations for attachment phenomena can be integrated with newly proposed information processing structures and mechanisms - such at those suggested by the Extended Mind framework [12].

3 THEORETICAL FOUNDATIONS OF ATTACHMENT THEORY

3.1 Ethological underpinnings

Attachment Theory is an evolutionary theory which was founded upon the Behaviour System concept from Ethology. In this framework Behaviour Systems control behaviours such as mating, fighting and feeding. Each Behaviour System carries out a species specific function, and has survived in the genome because its functions contribute to biological fitness. Therefore Behaviour Systems are related to one meaning of the term 'instinct' [10]. The Behaviour Systems that Bowlby linked to attachment behaviour in human infants are the attachment, fear, sociability and exploration systems [2]. The behaviour systems most closely related to attachment are inherently motivated [9]. This means that infants will work to experience exploration, socialisation and security. These outcomes are therefore primary drives that are not activated as the by-product of any more fundamental process.

Behaviour Systems within the Attachment Control System allow a flexible repertoire of behaviours to be produced when pursuing currently active goals. What defines the attachment control system and its constituent behaviours is the outcomes that reliably follow from activating these behaviours. For example, if an infant is anxious and its current goal is to increase its proximity to a carer the infant may cry (which predictably brings the carer closer), or crawl towards the carer themselves. Of course, crying and crawling may also be activated by other behaviour systems, such as, respectively hunger or exploration.

For Bowlby, behaviours resulting from the attachment behaviour system and the fear system have the predictable outcome of maintaining access and proximity to its primary carer. They might also involve facilitating access to the benefits of cognitive extension.

3.2 The attachment control system involves a hierarchy of forms of information processing

Attachment behaviours can be observed from infancy to adulthood. In adulthood contexts range from caregiving to romantic relationships. So it is unsurprising that a diverse range of information processing structures and mechanisms have been invoked to explain the diversity of empirical findings. From Ethology, Bowlby introduced relatively more simple concepts and mechanisms. In addition to the Behaviour Systems concept mentioned in section 3.1, other ethological concepts drawn into Attachment Theory include Reflex Actions and Fixed Action Patterns. Bowlby also showed how these 'simple' mechanisms can interact in complex ways by chaining and alternation. Behavioural patterns arising when fixed action patterns have been trained into complex sequences can be mistaken for behaviours directed by more complex goal corrected mechanisms because of the sensitive matching of response to stimuli.

Goal corrected feedback mechanisms were also introduced by Bowlby from control systems theory and cybernetics to play an important part in the 'purposiveness' of the attachment control system framework. Whereas younger infants may produce complex behaviour by chaining and training of simple mechanisms like reflexes and fixed action patterns, older infants, children and adults use more complex control mechanisms. Simple and complex mechanisms coexist, with each sometimes overridden by the other. So higher level processing afforded by cognitive extension should integrate with lower level processing.

Bowlby invoked internal working models (IMWs) and plans as explicit internal representations in the attachment control system. The internal working models (IWM's) concept was used to represent models of self and other in attachment relationships. In this role IWM's do not capture every aspect of reality but enough that the child can formulate plans and make decisions in relation to attachment goals. IWMs represent attachment related world knowledge and expectations about its caregiver's availability and responsiveness. These expectations are derived from the carer's past performance. According to Bowlby simple plans can be formed when several goal corrected steps are chained together, and each step must be completed before the next step is taken. More complex plans were also proposed where simple plans were formed into plan hierarchies. Ultimately plans come to be represented linguistically.

According to Bowlby, natural language is the ultimate and most sophisticated way in which individuals can represent themselves within their social environment. A benefit of the non-communicative aspect of language is that the possession of language allows more flexible and imaginative plans and subplans to be created. Bowlby

noted that linguistically formed representations also have the benefit that *"instead of each one of us having to build his environmental and organismic models entirely for himself, he can draw on models built by others"* [2, page 82].

This paper is thus an attempt to integrate Bowlby's idea that we can rely upon models built by others with the conceptual framework provided by the extended mind concept. In this integrated view, we not only reply upon models built by others, but these models are also held 'externally' and we access these models through our attachment figures. Sometimes, extended cognition may occur in attachment interactions in an enactive manner. For example, in infant can be viewed as possessing embodied enactive models of physical holding and will therefore expect to 'sink in' to a greater to lesser degree when being held according to their particular enactive model built from previous experience. However, Attachment Theory includes structural elements such as secure-base scripts which are not enactive but rather originate from a more traditional computational and representational approach within Cognitive Science. Attachment Theory therefore provides an explanatory framework where these diverse approaches are intimately and naturally linked. So one contribution of developing an extended mind approach to attachment phenomena is it provides a canonical example of social extension which also integrates a traditional computational and representational approach to explaining behaviour with enactive explanatory elements.

According to Bowlby the attachment control system develops from being reliant on simple mechanisms such as reflexes through many intermediate forms to finally being mediated by complex high level representations such as natural language. During this process of change there is often stability in the individual differences in behavioural patterns exhibited by individuals. This continuity is in part explained because often early appearing fragments of instinctive behaviour are integrated into later appearing complete sequences with their normal mature functional consequences. Cognitive extension is an additional cause of stability over time.

These developmental processes described by Bowlby involve intimate interaction between lower level processes, such as simple reflexive responses, and emerging higher level structures and mechanisms. New resources are created over time. Integrating elements into a system depends on (i) biases in infant learning abilities and (ii) information/structure in the expectable caregiving environment. The control system as proposed by Bowlby is not just preformed and waiting to be triggered or maturing without experience, but its rather constructed - through interaction between infant learning abilities and information available in the structure of the caregiving environment. This theoretical approach provides a ready framework to fit with extended mind mechanisms. A cognitively extending carer can provide access to resources which mesh with those already possessed internally.

4 ATTACHMENT THEORY - FROM PHYSICAL PROXIMITY TO INFORMATION PROVISION

Attachment Theory describes how our closest relationships develop and function across the life span. Attachment bonds are formed early in infancy and can reform and develop through the life-span. The hallmarks of attachment include (a) preference, (b) familiarity, (c) relative uniqueness, (d) identity, (d) use as a secure base, and (e) grief and mourning in response to loss [13]. So typically, only a very small number of attachment bonds are formed, most often with primary caregivers and, later, partners in enduring relationships and one's offspring. In infancy, as in adulthood, we tend to form attachment relationships with only one or a few figures at a time. Any of these may serve as a secure base from which to explore and a haven of safety. Early in life secure-base behaviour may be observed when infants attempt to gain and retain physical proximity with their attachment figures. Later in life adolescents and adults may retreat to their attachment figures without gaining physical proximity by using communications technology to make emotional contact.

Whilst Bowlby was setting out the information processing underpinnings for Attachment Theory, Mary Ainsworth and co-workers [1] studied how differences in infant-care interactions can affect the course of emotional and social development and the growth of attachment. The focus on individual differences in attachment status and development led to an empirically productive new direction for attachment research. Much of the contemporary attachment research on mental health issues and psychopathology is linked to individual differences attachment categories derived from the Strange Situation Experiment [1]. This is not an experiment where infant-carer dyads are randomly assigned to different conditions in the laboratory. Rather it is a standardised laboratory procedure where all infants are presented with the same controlled and replicable set of experiences.

To capture infant responses to changes in context, the Strange Situation procedure consists of 8 three minutes episodes which are designed to activate, intensify or relax the one-year-olds attachment behaviour in a moderate and controlled manner. The infant and carer enter the laboratory setting together, but then undergo a separation. The carer leaves from the room, before a reunion in a subsequent episode. As the first reunion episode ends the infant meets an unfamiliar stranger in the laboratory, before a further separation. In each episode infant behaviour is carefully recorded from behind a two-way mirror. In the final episode the mother is reunited with her one-year-old infant after the infant has been left alone for three minutes in the unfamiliar setting.

The infants responses to context changes that occur in the transitions between the eight episodes demonstrate typical normative trends. For example, infants, irrespective of home environment, typically exhibit increased distress when their carer leaves the room so the infant is left with a stranger (in episode four) or completely alone (in episode six). Nested within the normative trends are several patterns of response reflecting the infants confidence in the caregivers response patterns. The infants response in the reunion episodes correlates strongly with patterns of maternal behaviour and infant responses intensively observed throughout the previous year. Therefore a key finding of the Strange Situation, and which makes it such a valuable research tool, is that infant behavioural patterns observed when the carer returns to the infant after a separation (infant-carer reunions occur in episodes five and eight) provide the best short-hand classification for the attachment behavioural patterns of infant and carer observed at length in the home environment (1).

Individual differences in the Strange Situation cluster into four patterns:

- Secure (type B) infants form the largest proportion in non-clinical populations and secure behaviour is the reference pattern against which the other classifications are evaluated. Infant responses in reunion episodes in the Strange Situation include approaching their mothers in a positive manner and then returning to play and exploration in the room quickly. They receive care at home which is consistently sensitive, more emotionally expressive and provided less contact of an unpleasant nature; at home these infants are less angry and they cry less.
- Avoidant (type A) infants typically make up the second largest

proportion of non-clinical populations. Infants responses in reunion episodes in the Strange Situation include not seeking contact or avoiding their carers gaze or avoiding physical contact with her. These children return quickly to play and exploration but do so with less concentration than secure children. Whilst playing they stay close to and keep an eye on their carer. It may seem that they are not distressed or anxious in the Strange Situation. However, research employing telemetered autonomic data and salivary hormone assays has demonstrate that, despite their relative lack of crying, avoidant infants are at least as stressed by the procedure as secure and resistant infants. Their care at home is consistently less sensitive to infant signals and less skilled in holding the baby during routine care and interaction. At home these infants are more angry and cry more than secure infants.

- Ambivalent (type C) infants typically make up a small but measurable proportion of non-clinical populations. Infants responses in reunion episodes in the Strange Situation include not being comforted and being overly passive or showing anger towards their carers. These infants do not return quickly to exploration and play. Their care at home is less sensitive and particularly inconsistent. In comparison with average levels across all groups, C type carers are observed at home being more emotionally expressive; they provided physical contact which is unpleasant at a level intermediate between A and B carers and leave infants crying for longer durations. At home these infants are more angry and cry more than secure infants,
- Disorganised (type D). This last attachment pattern has been more recently categorised, is the least well characterised or understood, and forms a very small proportion of infants in the general population [8, page 26].

In a risk factors approach to psychopathology, secure status has been suggested as a protective factor whereas the three insecure attachment patterns have been suggested as risk factors for various subsequent psychopathologies.

5 ATTACHMENT, MIND EXTENSION, AND MENTAL HEALTH

Attachment Theory is a developmental theory but also an evolutionary theory. So we should expect that at every age, but especially in infancy and childhood, an effective attachment figure who is a primary caregiver should possess the goal to extend their cared for individual's mental capacities - elevating ongoing activities (enhancement) and supporting development to higher performance over time (enrichment). This might be taken to mean that an attachment figure aids those they care for by providing certain affordances in the social environment. This will be more important in infancy. In moment to moment interactions, a carer (if she recognizes what her infant is up to and is skilful and motivated) adapts her support (including extending and transforming it through the course of the interaction). She may reach for the approaching infant, then make adjustments as it comes close and makes efforts to be picked up. Much the same may occur in her support for exploration.

The Strange Situation emphasises that infants attempt to control physical proximity to their carers - by signalling to the carer or actually moving closer themselves when anxious. Or conversely moving further away to explore when feeling safe. More recent research has shown that young children do not just hold sensorimotor representations of the quality of their attachment relationships. For example, attachment relationship quality can be assessed by the simple drawings young children make of their families. This 'move to the level of representation' is well established in contemporary attachment research [8, page 36].

What is more novel in attachment research is a finding that infants and young children may filter information that they get from their carers according to how reliable the carer is in facilitating achievement of other goals that are not related to information gathering and exchange. A recent empirical result has shown that securely attached children hold a more nuanced and realistic view of their caregiver as a provider of information than insecure children do [6]. A traditional view in Attachment Theory is that attached children possess expectations about their carers as providers of physical and emotional care. This new result widens this perspective to see children as appraising the information providing qualities of their caregivers. So, in a traditional view, infants assess how effective their attachment figure is in manipulating the physical world. In this new view infants in addition assess how reliably carers process and provide information and may act to optimise how they can gain information from their carers. We can say that secure infants tend to use their carers more in 'epistemic actions' (where the action is not intended to change the state of the world but gain information about the world) [4]. Insecure infants not only trust their carers less to provide security, they also seem to trust their carers less to just provide accurate knowledge about the world [6]. So a secure 2-5 year old might use the mother as an extension of his/her mind. When asking for information he/she assumes that the mother knows what is being requested and assumes that the information provided is true and complete. Being egocentric at this age, the child would not distinguish between information sought from and provided by mother and information retrieved from his/her own memory. That is, limited source monitoring - knowing the answer and knowing that this is something mother told me are not well distinguished. Interaction with the mother facilitates development of a theory of their own mental processes and understanding others mental process. As Clark notes: *"The child is surrounded by exemplars of mind-reading in action, she is nudged by cultural interventions such as the use of simplified narratives, prompted by parental rehearsal of her own intentions, and provided with a rich palate of linguistic tools such as words for mental states"* [4, page 67].

There is substantial evidence that for an infant, gaining secure attachment status confers specific advantages in the subsequent development of social competencies [8, page 201]. Extended mind mechanisms such as use of carers in epistemic actions integrate well within a framework showing how secure attachment promotes social competence. When an infant is trusting of their carer they can use that attachment figure to extend their own mind as well as acting as a secure-base and haven of safety. This trusting relationship is likely to lead to less cognitive effort being needed by the infant when undertaking a range of tasks. So the extended mind mechanism provides part of a putative explanation for the upper reaches of cognitive performance. Therefore, attachment security and mechanisms of cognitive extension may be linked to good mental health in the sense of co-occurring with effective management of relationships and enabling general high-level cognitive performance.

Whilst the evidence linking secure attachment to improved social competence is relatively clear, the relationship of insecure attachment to mental health is not so straightforward. There are several possibilities for simple causal relationships between insecure attachment status and psychopathology which can be discounted. Firstly, insecure attachments are not a form of psychopathology warranting clinical attention, and are often adaptive responses to particular caregiving environments [8]. Secondly, empirical data show that insecure

attachment does not have a direct causal role in the later development of psychological disturbance. The effort to find the 'Holy Grail' of main effects of infant attachment on later psychopathology has so far been characterised as a *"fruitless search"* [7, page 638] Thirdly, insecure attachment is not even strongly linked to specific threats to mental well-being. Rather, there are multiple pathways to any given disorder. A single disorder might be reached from a combination of other risk factors. So in addition to early attachment relationships, other risk factors are: child characteristics such as temperament; parental socialisation and management strategies; and family ecology (family life stress and trauma, family resources and social support). Also, insecure attachment may contribute, along with these other risk factors, to multiple disorders [7].

A commonly accepted view is that early attachment is just one factor among many that either add protection or increase vulnerabilities to subsequent disorders. Attachment relationship dysfunction can give rise to serious psychopathology. For example, 'reactive attachment disorder' is one of a small number of psychopathological disorders diagnosable in young children [8, 7]. However, such psychopathologies are linked to significant abuse and negligence by carergivers. In contexts such as this, the presence or absence of extended mind interaction between an infant-carer dyad may not be a clinically useful measure. Nonetheless, Attachment Theory provides an evolutionary lense through which to view the socially-mediated development of mind extension. Unlike theories of cognitive scaffolding, Attachment Theory illustrates how mind extension can involve privileged access by an epistemically trusted carer to an infant. This privileged access may be part of the optimal evolutionary expectable caregiving environment. So possessing epistemic trust in a carer may be an important step or precursor to subsequent developments in many socio-emotional developmental trajectories towards mental health. Therefore, failure to experience such epistemic trust may be a contributory factor, rather than on its own push an infant into psychopathology. In addition, the multiple root causes of key psychopathology symptoms may interfere with interactions and learning necessary to develop confidence in a caregiver's availability (and hence extended mind capacity). Thus these multiple root causes and secondary and tertiary effects arising from attachment or extended mind interactions may play a role in the development and tuning of each other.

6 CONCLUSION

Attachment figures are situated centrally in complex interactions. Attachment Theory has set out details of an infant's information processing infrastructure which supports infant cognition in these contexts. Up till now Attachment Theory has focused on representations such as Internal Working Models and plans which are held 'within' the infants head. In this type of tightly coupled interaction, cognitive processes can bridge the traditionally conceived boundaries between social actors. The extended mind thesis suggests that computational processes:

> *"do not properly decompose into a neat sphere of internally achieved computations surrounded by a well behaved nimbus of calls to the world. In place of such a neat "inside-outside" boundary respecting cycle, we confront a bunch of unfolding internal processes, each of which is directly issuing at different timescales, calls both to other inner processes and to outward-looping epistemic acts that result in cognitively crucial episodes of closed-loop interaction"* [4, page 73].

The thesis put forward in this paper is that some carers may be considered 'transparent equipment' within the cognitive processes of infants with whom they are attached [4, page 80]. Not every child-carer dyad can be expected to develop mind extension. The child needs to trust their carer epistemically as well as a secure-base and haven of safety.

We cannot know what the two originators of Attachment Theory would make of the extended mind approach to Attachment Theory presented above. It is possible that John Bowlby would see parallels to psychoanalytic ideas on the absence of self-other distinction and its early development. Mary Ainsworth had a much more informed and finely tuned sense of mother-infant interaction. She may have recognised that her maternal sensitivity scales, particularly 'sensitivity to signals', and 'cooperation vs. interference with ongoing behaviour', [1] include similar ideas to the extended mind approach to attachment presented above. This similarity means that the extended mind concept is likely a useful tool for highlighting the level at which infant and maternal behaviour are coordinated.

Further research may explore how children develop from implicit use of mother as mind extension to explicit use. Questions include whether there are required sequences in these kind of changes, as opposed to trajectories that depend on various facts about the different genomes or different environments. In addition, work should also clarify how an extended mind approach integrated within Attachment Theory contrasts with other developmental approaches such as those that rely on cognitive scaffolding without affective support; and those concerned with mind-mindedness.

ACKNOWLEDGEMENTS

We would like to thank the referees for their comments which helped improve this paper.

REFERENCES

[1] M. Ainsworth, M. Blehar, E. Waters, and S. Wall, *Patterns of Attachment: a psychological study of the strange situation*, Erlbaum, Hillsdale, NJ, 1978.
[2] J. Bowlby, *Attachment and loss: volume 1 attachment*, Basic books, New York, 1969. (Second edition 1982).
[3] M. Carpenter, K. Nagell, and M. Tomasello, 'Social cognition, joint attention, and communicative competence from 9 to 15 months of age.', *Monographs of the Society for Research in Child Development*, **63**, (1998).
[4] A. Clark, *Supersizing the Mind: Embodiment, Action, and Cognitive Extension*, Oxford University Press,, New York, 2008.
[5] A. Clark and D. Chalmers, 'The extended mind', *Analysis*, **58**, 10–23, (1998).
[6] K.H. Corriveau, P.L. Harris, E. Meins, C. Ferneyhough, B. Arnott, L. Elliott, B Liddle, A. Hearn, L. Vittorini, and M.S. de Rosnay, 'Young childrens trust in their mothers claims: Longitudinal links with attachment security in infancy', *Child Development*, **80**(3), 750–761, (2009).
[7] M. Deklyen and M.T. Greenberg, 'Attachment and psychopathology in childhood.', in Handbook of Attachment: : Theory, research, and clinical applications*, 2nd Edition, eds. J. Cassidy & P.R. Shaver*, 637–665, Guilford Press, London, (2008).
[8] S. Goldberg, *Attachment and Development*, Arnold, London, 2000.
[9] 'H. Harlow, 'The Nature of Love', *American Psychologist*, **13**, 573–685, (1958).
[10] R.A. Hinde, 'Ethology and child development', in *Hand book of child psychology, eds. P.H. Mussen*, 27–93, J. Wiley and Sons, New York, (1983).
[11] D. Petters, *Designing Agents to Understand Infants*, Ph.D. dissertation, School of Computer Science, The University of Birmingham, 2006. (Available online at http://www.cs.bham.ac.uk/research/cogaff/).

[12] D. Petters and E. Waters, 'A.I., Attachment Theory, and Simulating Secure Base Behaviour: Dr. Bowlby meet the Reverend Bayes', in *Proceedings of the International Symposium on 'AI-Inspired Biology', AISB Convention 2010*, 51–58, AISB Press, University of Sussex, Brighton, (2010).
[13] D. Petters, E. Waters, and F. Schönbrodt, 'Strange carers: Robots as attachment figures and aids to parenting', *Interaction Studies: Social Behaviour and Communication in Biological and Artificial Systems*, **11**(2), 246–252, (2010).
[14] D.W. Winnicott, 'The theory of the parent-infant relationship', in The Maturational Processes and the Facilitating Environment*, eds. D.W. Winnicot*, 37–55, Hogarth, London, (1965).

In the Quest of an Objective Criteria for Defining Mental Disorder: an Evaluation of the Framework Comprised by the Systemic Analysis of Functions and the Extended Mind Hypothesis

Mariana Salcedo Gómez[1]

Abstract: The etiological theory of functions is the framework in which Wakefield develops his harmful dysfunction analysis of mental disorder. The purpose of this paper is double: first, object the role that plays the etiological theory of function as a suitable framework to account for disfunction in mental disorder, and, second, suggest that the systemic analysis of function and the extended mind hypothesis, could accomplish a better role in the definition of mental disorders as harmful dysfunctions.

1 INTRODUCTION

The etiological theory of functions is the framework in which Wakefield [1] supports his definition of mental disorder as harmful dysfunction (HD).Wakefield's proposal is a critical and purposeful answer to the first definition included in the third edition of the Diagnostic and Statistical Manual of Mental Disorder (DSM-III-R) published in 1987 [2]. This first official definition was qualified as operational given the explicit purpose of Robert Spitzer [3] and his colleagues of avoiding any commitment to any theoretical framework, and with the intention to fulfill the demands of several groups that objected the scientific character and objectivity of some psychiatric categories. Particularly, a central issue that Wakefield [1] observed in this operational definition was the lack of an objective criterion for specifying when we are facing "a manifestation of a behavioral, psychological, or biological dysfunction in the person", and when we are facing another undesirable condition that causes, as mental disorders do, unexpectable distress or disability to the person. Though Wakefield considers that this operational definition in DSM-III-TR is similar to his harmful dysfunction approach, given that contains at least and ambiguous idea of dysfunction and an idea of harmful, this is not useful for discerning between mental diseases and other mental conditions that are not diseases. . In order to provide a framework for specifying what in biological terms could be judged as a dysfunction and what is not, and consequently discern for disease and not disease, Wakefield makes use of the notion of etiological function, so as to overcome any ambiguity.

If Wakefield´s proposal is so relevant to the discussion of mental disorder definition, it is because he reaffirms the need to gather, in a unified definition, the value and scientific components of this concept, and the need to clarify and to make a consensus, if that is possible, about the kind of biological mechanism that when suffers a failure produces the sings and symptoms of mental diseases. Along with Boorse' notion of disease [4], in Wakefield's account of mental disorder lies the notion of function as a theoretical category suitable to medical field which enables to divide a system into smaller and simple parts, identifying how every part must work, the way in which it could fail and the plausible consequences of this failure. Wakefield, in particular, departs from an evolutionary perspective of biological functions i.e. , functions that refer to antecedent causes that result in a specific biological trait within an ancestral population. This evolutionary perspective is also known as the etiological theory of function attribution, and is a naturalized answer to the teleological meaning that historically was given to the idea of purpose that justified the existence of live organisms and human artifacts.

This etiological theory was a result of a modern history approach to functions [5] that emerges in the context of the unificationist program of science in the decade of 1970, when there was a renewed interest in philosophy of biology to recover the notion of function in order to maintain the explanatory autonomy from physics. In this scenario, two main naturalized programs of the notion of function were developed: the etiological theory, mentioned before, and the systemic analysis perspective. The etiological theory was developed initially by Larry Wright [6] in his article of 1973, *Functions*, an effort to which Ruth Millikan [7] added in 1989 with her work *In Defense of Proper Functions.* The systemic perspective on functions was authored by Robert Cummins [8], who, in 1975 wrote his *Functional Analysis* where he objects, in first instance, the unnecessary reference to the past in order to attribute function to a trait. Cummins instead pointing out that a systemic analysis of functions implies (y) referring to what a biological trait does, how it does it, and what its contribution is to the whole system, and not to the causes of its being there.

In the context of these two main theories of function attribution, the etiological and systemic view, the purpose of this paper is double, (1) object the suitability of etiological theory of function attribution as a framework to account for mental disorder, what means, object the harmful dysfunction analysis of mental disorders, and (2) evaluate the plausibility of a broader approach comprised by systemic analysis of function and extended mind hypothesis. My objection to the etiological theory of function attribution has a double via. In

1 PhD Student, National Autonomous University of Mexico, México, sagm75@gmail.com

first instance, I object the constraints that etiological theory imposes for attributing function to mental mechanism, given the natural selection history that must be appealed in order to characterize functionally any biological trait. In second instance, and as a consequence of the first objection, what we have is a very narrow definition of mental disorders, i.e., of the kind of conditions over which we can "justifiably" judge there is a "real failure" of a "real functional mental mechanism". In order to overcome the objections to harmful dysfunction analysis, at the same time that preserve an objective criteria to define mental disorder, I shall to argue that the systemic analysis of function proposed by Cummins [8,9], and the hypothesis of extended mind developed particularly by Clark and Chalmers [10], could be useful in the achievement of this enterprise. About my second purpose, I shall to show the qualities of the systemic analysis of functions in order to suggest that this is a more accurate framework, to better account for the sort of functional mechanism of mind and behavior, and consequently, a suitable approach for identifying failures that could arise in mental, cognitive or behavioral systems. I suggest that the unit of analysis of extended mind hypothesis could fits in a better way with the kind of systems we need to bound in order to identify and explain mental dysfunctions.

The proposal of considering the systemic analysis of function and the extended mind hypothesis, as part of an objective criteria in the definition of mental disorder, could be a solution to the problem of narrowness in HD analysis of Wakefield. One, because this approach at the same time that wide the conception of mind and of mental mechanisms, also wide the criteria of function attribution in the context of a broad mental system that goes beyond the brain/mind conception, and left apart the condition of natural selection and adaptive traits, as the only once that are subject of functional characterization.

2 BIOLOGICAL CRITERIA OF DISEASE AND ETIOLOGICAL THEORY OF FUNCTION IN WAKEFIELD'S DEFINITION

Scadding [11], Kendell [12] and Boorse [4] were the first to introduce a biological perspective of function in order to establish an objective criterion for demarcation between diseases and other medical conditions. Scadding [13], more than being interested in proposing a specific demarcation criterion in biological terms, his main goal was to make a broad characterization of diseases as "biological disadvantages", considering that a very precise biological notion was an useless enterprise in the context of such an heterogeneous spectra of diseases causes –bacteria, virus, malformations, tumors, lesions-, and the irregular character of empirical evidence that defines them: clinical evidence, anatomical or physiological dysfunctions, or causes of disease. A common problem observed by Scadding [11] was the confusion between the name and definition of diseases with its causes, instead of make the right identification, for example, between clinical description or the underlying anatomical or physiological disorder. The conviction of Scadding [11] was that the improvement of medical science will be directed through the search for etiology, which means that in the future, neither clinical characteristics, nor functional problems will be predominant in the definition of diseases. -even though some non-etiological diagnostic could coexist along with predominant etiological diagnostics.

In order to precise and clarify the notion of biological disadvantage, Kendell [42] and Boorse [4] proposed, independently, a pair of consequences contrary to the nature and survival of species that could be derived in the case an individual is in a situation of biological disadvantage: fertility reduction and decrease of life expectancy. Boorse [4] goes further in the characterization of a biological disadvantage and translates this notion as a function deficit. What this means, is a lack or a diminished capacity of an organ or system to perform that for which it was designed and that contributes to the survival and reproduction of the whole organism. Therefore, Boorse [4] defines disease as a kind of internal state that injures or harms health, and consequently reduces one or more functional capacities below its typical efficiency (1977:555). Boorse [14] adopts a biostatistical approach of functions. What this means is that every functional deficit is measured from the standard causal contribution of a trait to accomplish the organism objectives, that is to say, to survive and reproduce. Scadding [13] considers Boorse's approach [14] mistakenly focused in a functional deficit notion, ignoring or neglecting the variety of existing deficits in the medical gnosology, whereas his general notion of biological disadvantage pretends to recover a natural sense of abnormality that could be present under different kinds of injures in disease definitions. Nevertheless, both Scadding [11] and Boorse [4], postulate as consequences of biological disadvantage the same results, lowered survival and lowered reproductive fitness, establishing with this a characterization of the particular kind of harm implied when the natural order of an organism is altered. In this point, we find a central difference between Boorse [4] and Wakefield account of evolutionary notion of function. While Boorse [4] considers the harm caused by dysfunction in terms of reproduction fitness and survival, i.e., within what he considers biological values, Wakefield's hybrid definition of disorder sets aside the harmful criterion in a normative level. Wakefield argues that lowered survival or lowered reproductive fitness do not cause a real harm in the well-being of individuals, at least not in an individual scale that is relevant to diseases, even when from an evolutionary scale and at the level of species could cause an important decrease in the population rates ([1]:379). Consequently, Wakefield introduces a criterion of harm, independent of evolutionary theory he adopts, and in that sense, claims that a complete account of disease must include a double component, a valuable and an objective one, proposing his harmful dysfunction analysis of mental disorder.

From these two criteria, he concludes, is possible to claim that a person is diseased if and only if there is a failure in the way a mechanism was designed to perform, and that failure cause a "real harm" in the person affected. In his 1992 article, Wakefield claims that from an evolutionary account, it is possible to discern a simple biological variation from a lesion that is a disorder, because in the second case the trait couldn't perform its capacity as in the original way for which it was designed, damaging the well being of the organism in a harmful way ([1]:375).

At first glance the proponents of the objectivist approach to disease seem to be interested in identifying, exclusively, the kind of biological processes underlying those conditions. However, the adoption of an evolutionary perspective on function/dysfunction category, particularly in Wakefield's proposal, was gaining a normative character, though naturalized, the etiological theory implies, in some sense, a "must be or do" of traits functionally characterized. What conferred this normative character to biological function in teleological theory, is the idea that the natural process of fitness and selection is the causal explanation of trait being there, and therefore the implication that the relevant effect for which it was selected, is the natural/normal performance of a trait, and any decrease, diminution, or complete failure, represents a problem insofar as normal functioning is the standard performance that conferred organism their adaptive advantage.

This naturalized/normative character of biological function category has a significant weight in the harmful dysfunction analysis of Wakefield, insofar as he seems to confer a strong ontological status to dysfunction category since his reference to "real harm" is: a *person is disordered only when some mechanism fails to perform the specific function it was designed to perform and the failure of the mechanism causes the person real harm."* ([1]:378). This quotation seems to suggest that a condition is a disease if and only if, a mental or physical condition is away from biological standards established by natural selection, and from cultural standard of well-being. Wakefield's hybrid definition of mental disorder, fuses the rules of nature and sociocultural norms, to give an account of a deviation that is harmful for an organism and for individuals, postulating both kinds of deviation as necessary and sufficient conditions for a disease.

3 OBJECTIONS TO HARMFUL DYSFUNCTION

Regardless of the pertinence of an evolutionary approach to define mental illness as a harmful dysfunction, it should be noted that Wakefield's definition is limited, even for those who hold an evolutionary one of disease, at least at the face of variety of hypotheses about physical and mental diseases that do not meet the harmful dysfunction criteria. An example is of this is Darwinian Medicine Program from which it has been postulated an hypothesis and an explanation for diseases, at least under five different models: dysfunctional model, adaptationist model, mismatch model, trade-off model and life-span model [15].

My objections against the harmful dysfunction analysis of mental disorder proposed by Wakefield, are mainly two:

I. The etiological perspective of function adopted as a theoretical framework for discussing the kind of mental mechanisms for which it possible to claim a failure, imposes, by the same nature of the final explanations it seeks to give, a very restrictive condition to the notion of mental disorder. The condition referred here is that a functional trait could be described like such, only in the case we could provide its history in terms of natural selection. What this means is that such trait had certain advantages over other traits, to increase the probability for the reproduction and survival of a population in ancestral times. This condition gives as a result, a category too narrow to account for the mechanisms that underly in mental disorders.

II. Even if the condition for identifying a mental mechanism in terms of adaptive traits were enough, there is one problem that remains in order to confirm an hypothesis of a mental mechanism functionally characterized: to trace back the relevant effects and the moment when these were selected. The objection about this is, if we cant get the appropriate empirical evidence to confirm the hypothesis that an adaptive perspective of mental mechanism demands

These objections could be dissected into three pieces in order to clarify the senses in which the proposed definition is narrow. A first sense in which the harmful dysfunction analysis of mental disorder could be considered narrow has been already mentioned by several authors -Cooper, [16], Murphy [17], Lilienfeld and Marino [18], and Adriaens and De Block [15]- who point out the fact that diseases in general include conditions that are not only dysfunctional, even in the evolutionary perspective. One example of this is the classification that Adriaens and De Block [15] make about models of diseases that Darwinian Medicine Program uses to accommodate the kinds of conditions that are considered in that way: dysfunctional model, adaptationist model, mismatch model, trade-off model and life-span model ([15]:7).

A second sense in which the HD analysis of mental disorder could be considered narrow, is derived from the assumption that only failures of mental mechanisms evolutionarily characterized, are "real" mental disorders. What this implies is that, if there is a mental mechanism without a history of natural selection, then, there is no mental disorder. Murphy [17] thinks that an evolutionary perspective of mental mechanism could play an heuristic role in the construction of an architecture of mind, but the picture of mind that this approach gives is yet very limited, and leaves aside too much of what, in fact, is part of the mind and of the disordered mind. As I understand this heuristic role, it is possible to think the mental mechanism postulated by evolutionary psychology as a basic or unmovable mental structure, around which we could encounter another mental mechanism/structures/processes not evolutionarily described, but relevant for the explanation of mind.

A third sense of narrowness that could be considered as a problem for mental disorder phenomena, and that is tightly related with second sense, is the inner model of mind adopted in evolutionary psychology -the theoretical perspective of the architecture of mind implied in Wakefield's account of mental disorder-. It could imply, for example, that the environment is only a resource of the information needed to be processed or a receptor of the data processed by the agent, and therefore could not be considered as a place where the dysfunction could be localized.

Cosmides and Tooby [19] in their work "Toward an Evolutionary Taxonomy of Treatable Conditions", acknowledge the scientific value of Wakefield definition, since is anchored to the idea of human architecture as a collection of functional units that can potentially have an impairment and to cause harm. Their own proposal of "treatable conditions" is based on a broad framework that considers harmful dysfunction as a piece of a greater whole, in which harmful

normative criterion is a constant that can result from a trait's normal functioning. Cosmides and Tobby's approach to treatable conditions, evidences independence of value judgments in relation to natural or biological standards, showing rather the multiple relations that could be observed between value judgments and functional/dysfunctional conditions. What they point as important from the evolutionary approach is individuation of functional units of mind, and indispensable framework since their point of view for understanding complexity of features and interactions involved in adaptive processes that give rise to a variety of conditions medically treatable.

"Is the mechanism in question operating in a way that produces the functional output it was designed to (i.e., was selected to)? is an answerable scientific question. Is the visual system recognizing objects? Are the incest avoidance mechanisms making the prospect of sex with family members distasteful? Can the person recognize that they have been cheated? To answer questions of functional integrity or impairment, one needs at a minimum to have correctly (a) individuated the mechanism from others, (b) identified and characterized its function (that is, the ancestral adaptive problem its design features were selected to solve), and (c) characterized its problem-solving design features and how they interact to produce the target set of functional outputs that would have increased fitness in ancestral environments" ([19]:457)

Following Cosmides and Tooby, even the Wakefield constrained definition of mental disorder, this is an important point or departure that highlights the necessity for a framework to construct an architecture of mind in terms of functional units, and a biological theory -in a broad sense- about mind diseases. As Murphy claim, an important problem that prevails in mental field is discontinuity and heterogeneity about what is mental and how the mental can get ill. An example of this is the visual system; it is a mental phenomenon for cognitive neurosciences, at same time that does not seems counterintuitive for psychiatry. Additionally, there is lack of consensus of the admissible ways in which mind gets diseased. But, even though it seems that a biological theory of diseased mind and its architecture necessary, it is questionable that an evolutionary approach of function, along its historical constrains, is the appropriate framework for psychiatry and sciences of mind in general.

It is a fact that in the last 30 years psychology, ethology, anthropology, and cognitive neurosciences have adopted an increasing interest in evolutionary approaches, searching for cognitive capacities that make Homo Sapiens an intelligent species, with cognitive, emotional and behavioral capacities that pave the way for an eminently social life. Studies and hypothesis in this field have been valuable to psychiatry undoubtedly, since these have provided novel explanations about the most representative mental disorder such as schizophrenia, sociopathy, autism, as problems related with impairment in social cognition ([20]:253). The growing interest in evolutionary theory and methodology, probably obeys to conviction that through evolution Homo sapiens has reached a complex cognitive architecture, with capacities as language, memory and consciousness [21] Questions about the functional heterogeneity of our nervous system, how we became the cognitive agents we today are, what cognitive mechanism configure human mind, and which kind of problem solving tasks are we designed for, have been a reason for adopting an evolutionary perspective as a way to look for ultimate causes.

The heuristic value of evolutionary approaches in mental sciences seems undeniable, judging from the proliferation of publications in this field. However, even when it seems promissory, the truth is that there is also great speculation, not only in the postulation of mental/brain mechanisms (modules) for almost for every single capacity, but also in the explanation of its adaptive advantages, other sort of problems attached to Wakefield's perspective of dysfunction.

In order to overcome the objections to etiological approach to function/dysfunction, and at the same time preserve an objective criteria to define mental disorder, I shall explore if a conflation of the systemic analysis of function, with the hypothesis of extended mind, could integrate a suitable framework for research and help to classify mental disorders. What I am trying to do with this theoretical and explanatory framework is to provide notions of function and of the mental that help to deal and to manage the complexity of mental disorders.

4 A BROADER APPROACH TO MENTAL DISORDERS: SYSTEMIC ANALYSIS OF FUNCTIONS AND EXTENDED MIND HYPOTHESIS

In order to overcome the objection of narrowness that harmful dysfunction analysis inherits, I shall try to sketch a broader account of mental disorder that at least satisfies two conditions:

i. To be broad enough to admit as mental disorders, conditions that are failure of internal mechanism of mind/brain structure, and failures at other levels of explanation that could have a localization across the system of extended mind. This implies that our theoretical framework of mental disorders will admit not only failures in internal mental mechanism that could be evolutionarily traits, but also, failures of processes that are not necessarily attached to an identified evolutionary mechanism, but that are nevertheless elements of the mental, and therefore, of a disordered mind.

ii. To be constrained enough that the failure identified is related to the subject and his/her mental life and subjective experience. With this second condition, the pretension is to guarantee that the harm or impairment concerns the subject, and as the definition of Cooper ([16]:22) points out, the condition is a bad thing to have and the afflicted person is unlucky.

The idea that functional traits are those that conferred an adaptive advantage to a population in ancestral environments [7], seems to be partially relevant to the context of mental disorder explanations. But the concepts of disease and health could not be defined exclusively in terms of evolutionary standards. That is the reason why we need to search for a perspective of function that embrace biological standards in a

broader sense. In this sense, a plausible strategy is to adopt the systemic account of function proposed by Cummins [8], whose main goal is to describe how in fact sciences uses functional language[2].

Cummins's goal was to explain functional ascriptions from the perspective of a particular system and its performance. In this particular sense, a functional explanation of the kind "X have the function of doing F" makes reference to, according to Cummins, the way in which X contributes causally to maintenance and performance of a system S when X executes the function F. Here it is worth to point out, that system S is determined by research interests, and, consequently function F is dependent to selection of S. This dependence makes the concept of function an instrumental one, in the sense that is an analytic tool for the researcher, without the strong ontological compromises of an etiological theory.

In order to extend my description of functional analysis, it could be said that: (a) given a system S, the researcher decomposes it in pieces with established relations; (b) the researcher determines the causal interactions of X with other components, necessary for the successful performance of S; and (c) the researcher identifies the disposition Y of the trait X that is implied in causal interactions, such that the disposition Y is considered as function F.

The decomposition of the causally connected pieces that is referred in (a), corresponds to the structure of the system S. A successful functional analysis is one that meets the condition of establishing a structure at the right depth at the level of analysis, at the same time that it allows to maintain the structure regardless of the level of analysis required. In order to clarify the idea about "depth level of analysis", we can use the example of the human body as a unit of analysis. In accordance with certain goals of study, we can dissect the body into subsystems: respiratory, digestive, circulatory, etc. But the nature of the study could lead us to increase the depth of analysis, taking one of these subsystems, for example, the circulatory system, and decompose it into new elements, as might be, for the circulatory system, the heart or the various blood vessels, each of which would have a particular function. A deeper analysis, could choose one of these elements, for example the heart, distinguish each of its parts, such as valves or ventricles, and identify the capacities that each carries out, in an orchestrated way, so as to let the whole system accomplish its major or more complex function. A functional analysis explains the major function of a system as the programmed execution of the capacities of each component, and as in an assembly line where each simple component accomplishes its capacity because of the accomplishment of the other simple components. An important characteristic of systemic analysis is that capacities of each component help define the capacities of the other components. This last thing is what allows the possibility for an overall explanation of processes and constituents by knowing schematically the structure and interactions of a system.

One of the virtues of systemic view for psychiatry, is the possibility to individuate mental systems and create strong correlations between components, grounding this work in a systematic research in molecular biology, genetics, neurobiology, cognitive neurosciences, cognitive psychology, evolutionary psychology and psychiatry. An assumption that underlies my conviction that the systemic analysis of function is the best path for a definition of mental disorder, is that causal correlations of these phenomena are not linear or unidirectional, instead they establish causal correlations in different directions and at different levels. The suitability of systemic analysis rests in the explanatory strategy that characterizes this approach, and this consists in cutting the whole on simpler pieces in order to analyze the complexity of a system. Such explanatory strategy also includes a decomposition and identification of each piece at different levels of analysis.

Once we have a brief characterization of systemic approach to functional analysis, it would be pertinent to analyze the relevance of substituting the traditional perspective of mind as computational processes that happen inside the agent, with the extended mind hypothesis proposed by Clark and Chalmers in 1998 [10]. This hypothesis questions the traditional idea about the nature of mind and the study of cognition, by posing the question "where does the mind stop and the rest of the world begin?". Fundamentally, the extended mind hypothesis posits that a more appropriate unit of analysis for understanding the mental, is beyond the skull, extending the boundaries to the material and social context in which cognitive agents are involved. Clark and Chalmers postulate a unit of analysis for cognitive processes that is labeled *coupled system*, and whose principal characteristic is its two-way interactions between human organism and external world, and that this trait enables it to constitute a "cognitive system in its own right".

This active externalist account of mind adopts a functionalist approach, giving equal causal roles to internal and external vehicles of cognitive processes; attributing the same power for directing behavior to both internal and external vehicles; and considering both indispensable for correct performance of cognition [22]. This means that human organisms, artifacts, and abstract norms, are all elements of same cognitive processes.

If material and cultural world, both external to agents, are considered as constitutive components of mental processes, and, if the weight of the hypothesis seems to rest in the assumption that all components of a coupled system have the same causal roles, then the important question to our subject is this: how could these assumptions play a relevant role in the explanation of mental disorders? Although initially the idea of an extended mind seems plausible as a comprehensive framework for individuate mental systems, the explanatory relevance for mental disorders fades if we ascribe the same weight to artifacts as part of our extended mind. Instead, it seems more convenient for our explanatory purposes, to focus in the sort of extension of cognitive processes related with the social and cultural scaffolding. This idea of cultural scaffolding was developed to support the notion of situated activity, developed by Hendriks-Jansen [23], when considering the way in which early children acquire mastery for tasks as looking, pointing, reaching, grasping, and other

2 Millikan (1999) in her article *Bio-functions* is convinced that proper functions and systemic analysis of functions are not very different, and that proper functions are intrinsically attached to a systemic perspective. Might be worthy to review closer hers arguments, in order to decide if both perspective are in fact exclusive.

motor and intentional skills that are achieved in a systematic way inside "a natural environment of purpose-built artifacts and adults who think in intentional terms".

Even when this idea of cultural scaffolding is not derived straightforwardly from the hypothesis of Clark and Chalmers [10], it is clearly a notion inherited from the work of Clark and Karmiloff-Smith in the early nineties, based on studies in developmental processes in young children. What this notion highlights is that every cognitive capacity for acquiring, learning, developing, etc., is possible only because of the close interaction and manipulation of our entire organism -brain/body- with a natural and cultural environment. The complete picture of an extended mind includes, both, the facts present from the day we were born and our engagements with the world. As Menary ([22]:229) points out, "our engagements with the world are embodied...they are primary sensory motor engagements; when adults these embodied engagements remain in the form of manipulation of environment; hence some cognitive processes are hybrid processes...they comprise neural processes and vehicles and bodily processes on environmental vehicles".

5 HOW SYSTEMIC ANALYSIS OF FUNCTION AND EXTENDED MIND HYPOTHESIS IMPROVES OUR UNDERSTANDING OF MENTAL DISORDER

In his work of 2006, *Psychiatry in the Scientific Image,* Murphy already proposed to discard the etiological perspective of function in favor of a systemic analysis to account for the type of dysfunction in the definition of mental disorders. At the same time, Murphy proposed to overcome the common sense intuitions about mind that prevail in Psychiatry. Even when his intention was not to completely discard the intuitions of folk psychology about mind, those that remain must be congruent with the principles of cognitive, neurobiological and social sciences. Since it is usually considered that folk psychology must not prevail over sciences about what mind is, in the same vein, Murphy rejects the common sense idea of disease as a condition that result from the dysfunction of an internal mechanism designed for a particular purpose. Such an idea of disease rests on two assumptions, the first linked with the old fashioned and limited conception of disease as a lesion of an internal tissue; and second, the idea that only those mechanism designed with a particular purpose are susceptible of being described functionally. In relation with this last assumption, Murphy [considers that sciences actually work in a different way in order to ascribe functions, from a mechanistic approach, that is, cutting in pieces and trying to explore the capacity of every piece and its contribution to the performance of a whole system.

The mechanistic approach that Murphy proposes as a starting point for thinking about the mind and the disordered mind, implies among other things, the admission that the failures or problems of the mind do not necessarily invoke broken internal mechanism. Or in other words, that not every mental disorder has an organic etiology. In this sense, Murphy proposes three main possible kinds of relations that give place to pathological behavior: (i) lesion or dysfunction in neuro-cognitive structures; (ii) dysfunction of neuro-computational structures, i.e dysfunction in learning processes and, (iii) normal performance of neuro-computational functions in an unfavorable environment. From this brief classification of the kinds of problems implied in mental disorders, Murphy wants to establish that not every mental disorder is a failure in the functioning of a brain mechanism, and that there are also problems of behavior in which there is no dysfunctional brain mechanism involved, instead there is "a combination of excessive or misdirected psychological drives, operant conditioning, and bad social learning" ([17]:72).

Murphy also noted that a distinction, as the one made by McHugh y Slavney [24], between concept of disease and concept of behavior, is irrelevant if we focus on the level of cognition. He suggested that psychiatry must adopt a neuro-computational perspective of mind in order to overcome the common sense intuitions of what counts as mental and what count as disease, and that it should support and ground the meaning of these categories in scientific evidence. What is remarkable of Murphy's proposal is that he, as others philosophers do, notice that is not possible to commit *a priori* with the claim that mental disorders are exclusively a lesion or a dysfunction in neuro-cognitive structures.

How to introduce a notion of dysfunction, so as to give an objective criterion for mental disorder, an avoid, at the same time, the assumption that such a dysfunction must imply a lesion o failure in an internal structure? The answer to this question is part of the undertaking of Murphy, and of my proposal here. The strategy for achieving this goal has been, as Murphy did, is to make use of systemic analysis of function. However, unlike Murphy, who still preserves a traditional notion about the unit of mind and of cognitive processes, here we introduce , the extended mind hypothesis. The conflation of both the systemic analysis of function and the extended mind hypothesis allows us overcome the narrowness of harmful dysfunction analysis of mental disorder and to avoid the objection about the search for appropriate empirical evidence to confirm the hypothesis that an adaptive perspective of mental mechanism requires. At the same time that the approach proposed guarantees or satisfies the conditions stated in the beginning of fourth section of this paper: to be broad enough to admit as mental disorders conditions that are failure of internal mechanism of mind/brain structure, and failures at other levels of explanation that could have a localization across a system of extended mind; and to be constrained enough so that the failure identified is related to the subject and his/her mental life and subjective experience.

The systemic analysis of function:

1. It overcomes the narrowness insofar as this analytic strategy allows to ascribe a functional description to a broad range of mental mechanism, inasmuch as such a trait not must fulfill the condition of being an adaptive one and to have a natural selection history.

2. As Cummins and Murphy point out, the kind of explanation that functional analysis gives us, is the kind of relevant explanation that talks about the contribution of a

mental traits to the performance of a broad system, at the same time that it tell us about the interactions between components of the system. And not only why the trait is there.

3. The methodological strategy of systemic analysis, consisting in cutting in pieces and identify the interactions in between, allows us to manage the complexity of the mental processes studied.

4. The extended mind hypothesis allows us to localize cognitive processes beyond the boundaries of the traditional cognitive agent, providing a broad framework for systemic analysis of functions. It also gives the opportunity to include in an strict sense, material and immaterial culture, and other cognitive agents, as part of the mental system of a given subject.

The sum of these two approaches allows us to maintain a dysfunction notion wide enough that it admits conditions beyond the inner structures of the traditional cognitive agent, and this include conditions involving dysfunction of some other element within our extended cognition unit. The fact that our unit of cognition extends beyond the brain and the processes taking place within the agent, permits us to justifiably postulate that a condition is a mental illness, even if there is no an injury to an internal brain structure. It suffices that some component of our system of extended cognition is operating in an "irregular" manner to affirm, within an objective and empirical frame, that the system in question suffers a cognitive dysfunction thereby causing a mental disorder. Such a notion of dysfunction, framed within the perspective proposed, allows us to keep an explanatory context inside the range of what is considered empirically valid.

REFERENCES

1. Wakefield, Jerome. (1992). “The concept of mental disorder, On the boundary between biological facts and social values”. *American Psychologist*, Vol. 47, No. 3: 373-388. American Psychological Association.
2. American Psychiatric Association. (1987). *Diagnostic and statistical manual of mental disorders* (3rd ed., rev.). Washington, DC: Author
3. Spitzer, R. L., & Williams, J. B. W. (1982). The definition and diagnosis of mental disorder. In W. R. Gove (Ed.), *Deviance and mental illness* (pp. 15-31). Beverly Hills, CA: Sage.
4. Boorse, Cristopher. (1975). “On the distinction between Disease and Illness”, *Philosophy of Public Affairs*, Vol. 5 No. 1, pp. 49-68.
5. Godfrey-Smith, Peter. (1984). "A Modern History Theory of Functions", Noûs, Vol. 28, No. 3 (Sep., 1994), pp. 344-362.
6. Wright, Larry. (1973). “Functions”, en *Philosophical Review*, 82: 139-68. *Philosophical Review*, 82: 139-68.
7. Millikan, Ruth G. (1989a). "In Defense of Proper Functions", Philosophy of Science 56.2: 288-302.
8. Cummins, Robert. (1999). “Functional Analysis”, en Buller, David, *Function, Selection and Design*, SUNY Press, Albany, NY.
9. Cummins, Robert & Michel Roth. (2010). “No traits have no evolved to function the way they do because of a past advantage”, en Robert Arp & Francisco Ayala, Eds., *Contemporary Debates in Philosophy of Biology*, Blackwell, pp. 72-86.
10. Clark, Andy & David Chalmers. (2010). "The Extended Mind". In Richard Menary Ed., *The Extended Mind,* The MIT Press, Cambridge, Massachusetts, pp. 27-42.
11. Scadding, J. G. (1967). Diagnosis: The clinician and the computer. *Lancet, 2,* 877-882.
12. Kendell, Robert E. (1975). The concept of disease and its implications for psychiatry. *British Journal of Psychiatry, 127,* 305-315.
13. Scadding, J. G. (1990). The semantic problem of psychiatry. *Psychological Medicine, 20,* 243-248
14. Boorse, Cristopher. (1977). “Health as a Theoretical Concept”, *Philosophy of Science*, Vol. 44, No. 4 (Dec., 1977), 542-573.
15. De Block, Andreas & Pieter Adriaens. (2011). “Why philosophers of psychiatry should care about evolutionary theory”, pp. 3-32. Adriaens, Pieter R & Andreas De Block, *Maladapting minds: Philosophy, Psychiatry, and Evolutionary Theory*, Oxford University Press.
16. Copper, Rachel. (2005). *Classifying Madness. A Philosophical Examination of the Diagnostic and Statistical Manual of Mental Disorders*. Netherlands: Springer.
17. Murphy, Dominic. (2006). *Psychiatry in the Scientific Image*. MIT Press, Cambridge, Massachusetts.
18. Scott O. Lilienfeld & Lori Marino. (1995). "Mental Disorder as a Roschian Concept: A Critique of Wakefield's "Harmful Dysfunction" Analysis". *Journal of Abnormal Psychology* 1995, Vol. 104, No. 3,411-420. American Psychological Association.
19. Cosmides, Leda & John Tooby. (1999). “Toward an Evolutionary Taxonomy of Treatable Conditions”, *Journal of Abnormal Psychology*, 1999. Vol. 108, No. 3, 453-464
20. Brüne, Martin, Hedda Ribbert, & Hedd Ribbert, Eds. (2003). *The Social Brain. Evolution and Psychopathology*. John Wiley & Sons, England.
21. Platek, Steven M, Julian Paul Keenan, & Todd K. Shackelford, (2007). *Evolutionary Cognitive Neuroscience.* The MIT Press, Cambridge, Massachusetts.
22. Menary, Richard. (2010). "Cognitive Integration and the Extended Mind". In Richard Menary Ed., *The Extended Mind,* The MIT Press, Cambridge, Massachusetts. 227-244.
23. Hendriks-Jansen, Horst. (1996). Catching Ourselves in the Act. Situated Activity, Interactive Emergence,

Evolution and Human Thought. The MIT Press, Cambridge, Massachusetts.

24. McHugh, Paul R., & Phillip R. Slavney. (1998). The Perspective of Psychiatry. John Hopkins University Press.

Hacking the extended mind?

Pete Faulconbridge[1]

Abstract. In lieu of the full, final paper, the following is an extended summary of the paper to be delivered. It constitutes a first attempt to think about the idea of 'hacking' a mind, surveying some recent examples of talk about this idea and suggesting that it may have some application in the context of the extended mind hypothesis.

1 HACKING THE EXTENDED MIND?

In the past few years, there has been a small but significant use of terms like 'mind hacking', 'self hacking', 'brain hacking' and 'person hacking'. These terms and variations have been applied to practices as diverse as meditation, hypnosis, perceptual illusions, the use of so-called 'smart drugs', mnemonic techniques, 'neuro-linguistic programming', the methods of standard cognitive science, diet and exercise regimes designed to optimise mental performance, CBT-influenced self-help, the human factors in network security and the use of brain-computer interfaces to gain access to personal information. The term 'hacking' here is generally used in what will be call the 'productive sense' (explained below) in which it refers, roughly, to a way of exploring and improving a system by practical investigation and adjustment. This is in contrast to the 'pejorative sense' under which it signifies unauthorised intrusion and manipulation of a system.

Whilst the term 'mind hacking' and related terms are not in popular usage, they have appeared in a number of books, magazines, blogs, interest groups and non-academic conferences.[2] I am generally sceptical of the notion of 'hacking' a mind, if taken literally, but I hope in this paper to provide some reasons to think that the concept might have some practical use in the context of the extended mind hypothesis and computer-facilitated cognitive enhancement.

There is no widely-accepted definition of a mind hack, and the term does not currently describe a single concept, practice or even set of interests. Perhaps, then, it is merely a metaphor which has caught the imagination of a few particular groups of people. Indeed, we may be sceptical of the very idea of hacking something like a person's mind. This paper is a first step towards evaluating some of the uses to which these terms have been put, and to explore the concept of hacking a mind, self, brain or person to try to determine whether it is simply a metaphor or whether it has some utility. The paper begins by surveying some interesting uses of these terms in recent media.

In order to understand the ways in which the term 'mind-hacking' is being used, it is necessary to understand what activities have previously constituted 'hacking'. The paper aims to elucidate the meaning of 'hacking' in the paradigmatic sense by exploring its origins.

A survey of the historical development of the concept will allow for a rough characterisation of paradigmatic 'hacking'. This practice seems to have emerged in MIT in the late 1950s as a response to certain types of highly complex, human-generated, significantly determinate (though not entirely predictable) systems (namely, the MIT Model Railway Club's switching and signalling system and academic computer systems). The paradigmatic 'hackerly approach' involves making use of knowledge of the design and 'rules' (in some sense) of the system, often in unforeseen or un-designed ways, in order to achieve a goal. A hack is rarely if ever based on the 'official' procedure, so hacking will often require trying out a given solution in practice, and its success will be hard to predict. Therein also lies the power of the hackerly approach – powerful and often elegant solutions can be found to problems which have no 'official' solution.

The above provides a rough characterisation of 'hacking' in what will be called the productive sense, in which it constitutes an established and effective approach to software programming, for example. It is suggested that the more well-known pejorative sense of the term, according to which it signifies malicious or unauthorised manipulation of computer systems, can be understood as an outgrowth of the productive sense. As mentioned above, although the term as used in the media commonly carries negative connotations, use of the term 'mind hacking' usually involves the productive sense.

With this characterisation in place, it is argued that the very idea of hacking a mind, self or person should be viewed with caution. There are philosophical arguments which suggest that changes to a system on the 'design level' are not apt for explanation on the intentional or personal levels, as would seem to be required for talk of 'mind hacking' to be apt.

Though we might object to such a strong autonomy thesis, this formulation suggests a practical constraint. Within the mind-as-information-processing-system model, it does not seem we currently have sufficient knowledge of the 'rules' of this system to allow for the development of a hacker-like approach to producing a substantial or interesting influence on the intentional behaviour of people.

Despite these problems with the use of this term, the paper goes on to suggest that, within the framework of Clark and Chalmers' extended mind hypothesis, there is conceptual and practical room for a practice which might be fruitfully thought of as 'hacking the extended mind'. According to this hypothesis, if an organism's thought or behaviour can be understood as arising from their forming the right sort of 'coupled system' with an 'external entity', this system "can be seen as a cognitive system in its own right." [5] It is proposed here that the context of two-

[1] MPhil student, Dept. of Philosophy, UCL, UK. Email: peter.faulconbridge12@ucl.ac.uk

[2] For some representative uses of such terms, see [1, 2, 3, 4]. Recent non-academic conferences include 'Self-hacking day event' (London, 2012) and 'Use Your Head: The Future of Mind Hacks' (San Diego, 2008) More examples will be discussed in the full paper.

way interaction between a computerised system and a person provides the most fruitful context in which to think about the idea of 'mind hacking'.

This broad claim is introduced through discussion of a recent example of 'hacking talk', and elaborated through a number of thought-experiments based on existing technologies. Given the historically context-insensitive nature of computing, extant examples of the right sort of coupling are hard to come by. However, it is suggested that it can already be seen in cases of, for example, prosthetic vision technologies and that the development of increasingly portable, context-sensitive and 'pervasive' computing will increase the scope considerably.

The second, shorter, part of the paper applies the insights of the first to topics of sensory and cognitive enhancement, and mental health. Whilst available technologies are generally directly implicated only in scaffolding relatively 'low-level' intentional behaviours, current work developing technologies to dynamically augment memory and attention, amongst others, promises to implicate human-technology coupling in areas much closer to meaning and rationality. Based on this prediction, some initial thoughts are offered as to how the benefits of hacking in the productive sense might be harnessed in development of these technologies, and how protection may be offered against the risks of hacking in the pejorative sense. It is suggested that mental health may be a particularly important area to think about both these benefits and risks. As we pursue technologically-mediated augmentation of our minds, our minds become more porous to technological influence, particularly when that augmentation is pursued in the context of existing vulnerability.

REFERENCES

[1] R. Hale-Evans. *Mind Performance Hacks: Tips & Tools for Overclocking your Brain*. O'Reilly Media (2006)

[2] J. Griggs. Brain-hacking art: Pictures that turn inside out. In: *New Scientist*, http://www.newscientist.com/article/mg20727781.000-brainhacking-art-pictures-that-turn-inside-out.html. (2010) (Accessed online 03/03/2013)

[3] G. Dayal. Researchers Hack Brainwaves to Reveal PINs, Other Personal Data. *Wired.com*, http://www.wired.com/threatlevel/2012/08/brainwave-hacking/ (Accessed online 03/03/2013)

[4] I. Mann. *Hacking the Human: Social Engineering Techniques and Security Countermeasures*. Gower Publishing Company (2008)

[5] A. Clark, C. Chalmers. The Extended Mind. *Analysis*, Vol. 58, No. 1, Jan. 1998, pp. 7-19. (1998)

Enkinaesthesia: Re-conceptualizing "Mental" Illness[1]

Susan Alice Jane Stuart[2]

1 ABSTRACT

The dynamic plenisentient[3] interrelation of agent and world is specified in kinaesthetic terms. Kinaesthetic activity, with its temporal-spatial-energic qualities[4], is always affectively-laden, and through the formation of intercorporeal resonances, the activity necessitates enkinaesthetic entwining with those agents with whom, and those objects with which, we are in relations of perpetual community. I will argue that the capacity for enkinaesthetic dialogue is an *a priori* nomological condition for agency and the generation of a felt anticipatory dynamics both within and between agents.

Enkinaesthesia emphasizes not just the neuromuscular dynamics of the agent, that is, the givenness and ownership of its experience but also the entwined, blended and situated co-affective feeling of the presence of the other[5] (agential and non-agential alike) and, where appropriate, the enkinaesthetically anticipated arc of the other's action or movement, including, again where appropriate, the other's intentionality. The 'other' can be sensing and experiencing agents and it is their affective intentional reciprocity which co-constitutes the conscious relation and the experientially recursive temporal dynamics that lead to the formation and maintenance of integral enkinaesthetic structures and melodies. Such deeply felt enkinaesthetic melodies emphasise the dialogical nature of the feeling of being as the feeling of being-*with* or being-*among*, and demonstrate the paucity of individuating notions that treat agents as singular.

Enkinaesthesia, as the openness to and reception of myriad subtle multi-drectional cues in dialogical relations, provides grounds for saying, following Heidegger, that it is this which constitutes the primordial mood of care[6] for human relationships and the deep roots of morality. If this is the case, then we might think of it as composing an 'ethiosphere' consistent with the semiosphere and the biosphere as presented by Hoffmeyer [28, 29].[7]

Finally, I will consider how "mental"[8] illness and, for example, grieving over the loss of a loved-one, can be reconceptualized in terms of enkinaesthetic fragmentation and failure.

2 INTRODUCTION

I will open this paper with a quotation from Evan Thompson's work and follow it with five preliminary theses which will be fleshed out in the body of the text. The quotation and each of the remarks should be used as a guide by the reader for the metaphysics of the terrain of ideas they are about to cross.

In the Preface to *Mind in Life: Biology, Phenomenology, and the Sciences of Mind* Thompson writes that the

> . . . incipient mind finds sentient expression in the self-organizing dynamics of action, perception, and emotion, as well as the self-moving flow of time consciousness.
> From this perspective, mental life is also bodily life and is situated in the world. The roots of mental life lie not simply in the brain, but ramify through the body and the environment. Our mental lives involve our body and the world beyond the surface membrane of our organism, and therefore cannot be reduced simply to brain processes in the head. [53, p.ix]

2.1 Five Preliminary Theses

- Boundaries are mutable and yielding.
- Consciousness and agency are co-constituting.
- Consciousness is the relata between agents and agents, and agents and objects.
- Causality is, at least, bi-directional, but more likely to be reciprocally recursive.
- The substance / state ontology is misconceived.

The boundaries which seem to separate us from our worlds open us up to those worlds and reveal to us our inseparability from them. Those boundaries which can appear, at first, rigid and fixed are often malleable and semi-permeable. We need think only of the skin with its surfaces within surfaces [29, pp17-38], the biological membranes of stratum corneum, epidermis, dermis and subcutaneous tissue, and our sense receptors and nerves; then there are the hairs that respond to temperature, which can stand erect if we are suddenly fearful, and which can be brushed by a sleeve or touched gently by a breeze; and then there are the non-biological membranes of clothes

---

[1] An earlier version of this paper appeared as "Enkinaesthesia, Biosemiotics and the Ethiosphere" in *Signifying Bodies: Biosemiosis, Interaction and Health*, (2010), pp.305–30, ISBN 978-972-697-191-7

[2] University of Glasgow, Scotland, email: Susan.Stuart@glasgow.ac.uk

[3] By 'plenisentient' is meant 'fully feeling' or fully switched-on sensory experience.

[4] These qualities are felt within an horizon of embodied, sentient activity and characterized by their directedness and energy. They are, I will claim, prenoetically intentional.

[5] 'Enkinaesthesia' is characterised by 'immanence', a term used by Deleuze and Guattari [16] to emphasise the direct, non-duality of the inescapable experience of 'other'. This is also emphasized in the use of 'enkinaesthesia' as opposed to 'interkinaesthesia' because (i) with the prefix 'en' the experiential entanglement of agent and agent, agent and object is emphasised, and (ii) it doesn't bastardize the Latin and Greek etymological roots.

[6] A "mood is primordial, meaning that it is presupposed by the intelligibility of all explicit forms of cognition and volition. It is a condition of sense for any encounter with beings, whether theoretical or practical." [40, p.239]

[7] The term 'ethiosphere' has a dual focus of derivation: firstly, 'ethi' has been taken from the term 'ethics', and secondly, since it is being developed within the contexts of the biological and semiological fields of enquiry it makes sense to speak of an 'ethiosphere', that is, the sphere of ethical enquiry.

[8] I have deliberately chosen to place "mental" in inverted commas because an unthinking adoption of the term implies an acceptance of the mental / physical dichotomy, and since enkinaesthesia is somatic and relational, the assumption of an ontological dualism would be false and unnecessary.

with their textures and degrees of translucency, and our personal and social boundaries which vary in relation to our moods and emotions, our confidence, our company, our feeling of well-being and health, and so on. Our natural assumption is to see the boundary of the body as the limit of our experiential world, but it is precisely its semi-permeable nature its breach which provides us with the possibility of experience in the first place. The skin, overrun with an abundance of receptors – sixty kilometres of nerve fibres, fifteen kilometres of veins, with millions of sense receptors for pain, temperature, pressure and touch [29, pp18] – opens us up to the world and discloses it through our inescapable engagement with it, and then, of course, the skin is supplemented by the plenisentience of visual, proprioceptive, kinaesthetic, auditory, gustatory, and olfactory senses which open us up in their own way, are affected by change or motion within our world and which, with internal feedback, can bring about affective change within themselves.

Agential bodies are co-affective sensory-kinaesthetic systems which spill out into the world and the lives of others. Embodiment may be a nomological condition for agency [19] but it is 'embodiment' broadly conceived, for it is the agent's capacity to transgress its boundaries, to spill over into the bodily experience of others, which establishes the community and reciprocity of felt co-engagement[9], and it is this felt co-engagement which is fleshed out in the expressive, meaningful and cognitive bodily dynamics which are, in themselves, the necessary precursor to effective affective social, cultural and linguistic communication in the human agent.

The living body or *Leib* isn't just given [24], that is, affectively present to the agent as part of its precognitive "operative intentionality" [35, p.xviii]. The living body in its reciprocally affective co-agential lived experience anticipates, imagines, and enacts how it expects prenoetically its world will continue to be. In so doing it is co-affective with other agents and objects, perpetually folding into, unfolding from, enfolding the other and *vice versa*, and all within an experiential temporal and processual horizon, not a series of punctuated and discrete temporal moments. It is within this "passive synthesis", where an agent is involuntarily affected and influenced by its world or *Lebenswelt*.[10] The necessarily relational co-agential reciprocity of this affection is 'enkinaesthetic', and is the manner in which we are open to the world.

'Enkinaesthesia' is a neologism I will use to refer to the reciprocally affective neuro-muscular dynamical flows and muscle tensions that are felt and enfolded between co-participating agents in dialogical relation with one another.[11] Enkinaesthesia, like intersubjectivity and intercorporeality relates to notions of affect, but in this case it is with the affect we have on the neuro-muscular dynamical flow and muscle tension of the other, including other animals, through our direct and our indirect touch. Direct touch includes the physical touch of a caress, a pat on the back, a hug, or the rebuff of the shrugged pulling away from contact. Indirect touch can be achieved through a look[12] where one becomes the object of someone else's subjective attention and experience, for example, in an unspoken admonishment, a papal blessing which can shrive us of your sins, a friend's wave from a departing train, or in the way words and language, as biodynamical engines[13], can alter the way we feel.[14]

3 THE FEELING AND SENSING BODY

The feeling and sensing body has gained prominence in discussions of consciousness and experience in recent years[15], including the work of Damasio [12, 13, 14]], Edelman [20, 21], and Sheets-Johnstone [43, 44, 45], and whilst I am generally sympathethic with these theories, they remain predominantly individual-centered and only minimally-interactivist in character. Noë's view [36, 37] comes closest to my own, moving away, though not entirely, from the self-centred view, though he remains a little shy of the full commitment I want to make to the enkinaesthetic reciprocal affective neuro-muscular dynamical flow that is felt between agents in dialogical relation with one another. Noë writes:

> The locus of consciousness is the dynamic life of the whole, environmentally plugged-in person or animal. Indeed, it is only when we take up this holistic perspective on the active life of the person or animal that we can begin to make sense of the brain's contribution to conscious experience. . . . Human experience is a dance that unfolds in the world and with others. You are not your brain. We are not locked up in a prison of our own ideas and sensations. The phenomenon of consciousness, like that of life itself, is a world-involving dynamic process. [37, p.xiii]

The moving, feeling, perceiving body is at the core of lived experience. But a non-relationally-situated sensory-kinaesthetics with little consideration of the affectively-laden interpersonal and interobjective world in which the agent finds itself will provide only a partial account of the experiential whole. Noë is right: the agent must be conceived from a holistic perspective, but the essential nature of the organism is not simply its kinaesthetic[16] unfolding in the world and with others; the holistic perspective must embrace the agent not simply as a being in the world but as, and always as, a being with the world, folding into, enfolding with, and unfolding from those other agents and things with which it co-exists *in utero* to the point at which we depart this life.[17] Ratcliffe [41] speaks of this experiential entanglement as phenomenologically primitive:

> World-experience is not distinct from how one's body feels; the two are utterly inextricable. The experiential entanglement of body and world is more phenomenologically primitive than experience of either in isolation from the other. [41, p.1]

[9] This notion of community and reciprocity is redolent of the A edition version of Kant's third Analogy where he states that "All substances, so far as they coexist, stand in thoroughgoing community that is, in mutual interaction." [32, A212]. No more can be said about this connection at this point but it is something that I intend to develop in future work.

[10] 'Passive synthesis' is a phrase used by Husserl to describe the pre- or non-linguistic, and thus, pre-conceptual sense-making which is the mark of our practical bodily, kinaesthetic engagement with our world. A simple example of this is the rhythm or momentum and muscular expectation we build up when walking on a moving walkway in an airport, or when walking down a moving, descending escalator. We make sense of how things are moving with us and we very quickly establish a kind of kinaesthetic prosody with our changing world. We move together fluidly until we leave the walkway or the escalator stops unexpectedly when we're forced to perform a more active, thought-full synthesis. [See Husserl [30].]

[11] By 'dialogical' I mean only the interactivity of agents and not textual, linguistic, or conversational activity.

[12] For an interesting elaboration of how we can be affected by the look of another read Chapter 1 of Part 3 of *Being and Nothingness* by Sartre.

[13] Stuart & Thibault "Enkinaesthetic polyphony as the underpinning for first-order languaging", [51].

[14] Direct touch may be straightforward to describe but experientially it is as vast and variable in effect and affect as indirect touch; the reason has to do with surfaces, boundaries, and borders, and what we perceive to be the limit of the bodily 'self'.

[15] For a nice summary of embodied cognition work, though with a little too much emphasis on language for my own taste, see [2].

[16] From here on 'sensory-kinaesthetics' will be encompassed in the term 'kinaesthetic'.

[17] For a commentary and discussion of enactive *in utero* development see Wood & Stuart [56].

Lived experience is, first and foremost, enkinaesthetic.

4 KINAESTHESIA AND THE PRIMACY OF MOVEMENT

So let's lay out the stall. The cognitivist view of the mind, that presents the mind as symbolic, representational, and reducible to a set of physical states and processes that are fully-explicable through scientific experiment and analysis, has been the predominant explanation for the mind in the second half of the twentieth century. At heart it is individual-centred[18] and utilises a substance-state ontology that treats temporality and spatiality as uniform, linear, and regular, consisting of discrete or punctuated events, points, objects, and places. On top of this it maintains the Enlightenment ideal of systematization – attempting to carve nature at its joints.[19]

Enactivism, on the other hand, emphasizes the agent's situation and embodiment in terms of its active, ongoing, processual, non-symbolic, non-representationally-based engagement in its world. It is essentially anti-dualistic, but unlike cognitivism's inclination towards a monist materialism, the enactivist ontological commitments are rather more complicated. The agent is embodied and dynamically-coupled to the world of other agents and things; thus, agent, world and action are necessarily intricately interwoven, and the agent's body, experience, action, and world together shape the way in which she deals with her everyday pragmatic concerns. Under this conception mind and world are inseparable, and it is embodied affective practice, rather than cognitive deliberation, that is the hallmark of the agent's engagement with her world. With only a slight modification enactivism embraces enkinaesthesia; the focal point moves from the agent and their individual agency to the necessity of our being co-agential in a co-dynamically continuous, affectively-laden intersubjective and enkinaesthetic processual horizon of experience. "By a 'way of finding oneself in the world"' Radcliffe says "I mean a sense of the reality of self and of world, which is inextricable from a changeable feeling of relatedness between body and world." [41, p.2] Thus it is that feeling bodies and things together in a dialogue of community and reciprocity with other feeling bodies and things play an integral role in full-bodied pre-linguistic sense-making relations.

> Babies in the womb... send and receive messages without benefit of the words, syllables, and phrases that begin appearing in a year or two after birth. Their daily experiences of communication are punctuated by self-initiated and reactive movements which express needs, interests, and feelings.... Based on the early development of the senses in the womb, a fetus remains in constant dialog with the surrounding environment. [5]

So, the genesis of this activity begins *in utero* and is necessarily co-agential, mother with prenate, occasionally mother with two or more prenates, and prenates with their bodies and the surrounding amniotic environment and beyond. "The maternal womb is an optimal, stimulating, interactive environment for human development. Activity never ceases and a fetus is never isolated." and, Chamberlain adds:

> Between week six and ten, fetal bodies burst into motion, achieving graceful, stretching, and rotational movements of the head, arms and legs. Hand to head, hand to face, hand to mouth movements, mouth opening, closing, and swallowing are all present at 10 weeks (Tajani and Ianniruberto, 1990). By 14 weeks, the complete repertoire of fetal movements seen throughout gestation are already in evidence (deVries, Visser, and Prechtl, 1985). Movement is spontaneous, endogenous, and typically cycles between activity and rest. Breathing movements and jaw movements have begun. Hands are busy interacting with other parts of the body and with the umbilical cord. From this early stage onward, movement is a primary activity, sometimes begun spontaneously, sometimes provoked by events. Spontaneous movement occurs earliest, probably expressing purely individual interests and needs. Evoked movement reflects sensitivity to the environment. For example, between 10 and 15 weeks g.a., when a mother laughs or coughs, her fetus moves within seconds. [6]

Our sensed and felt co-agency begins as soon as movement starts for this movement incorporates the sensations of touch, temperature, pain, hearing, balance and orientation, chemosensors of smell and taste, mouthing, and sucking and licking which are used to explore texture, hardness, and contours of objects, and, of course, the prenate's own body and, in the case of twins, the other's body too. Neither mouthing nor sucking and licking in this context are involved with eating and nutrition, rather they are, as are the others, affective dialogical means of exploration, and it is in this exploration, this non-propositional questioning of its world of felt-being-with, that the prenate is establishing its first field of values. Its *Umwelt* can be better described as its *Mitseinwelt*, the felt-being-with, for its experience is affectively-laden co-engagement; its touching, tasting, hearing is concernful exploration that matters and is values-realizing from the start.[20]

Through its enkinaesthetic sensitivity the agent can establish the reciprocal affective enfolding required for the timely response and adaptation it will need post-natally to survive, and the greatest advantage afforded the burgeoning agent is to feel as it moves, to move as it feels, and to begin to grasp its world *ab initio*.[21]

> We are deeply and naturally kinaesthetic and enkinaesthetic, aware of our bodily movement and our action in the world, but also able to affect others and be affected by them, moving and being moved [3] within a reciprocal affective neuromuscular dynamical temporal flow. The way in which these felt somatosensory relations fold and unfold by bringing forth our world through our kinaesthetic imagination and associated somatosensory expectations together influences how we will shape and adapt our world, how we will then adapt to those changes, and so on. [49, p.179-80]

Our unceasing kinaesthetic and enkinaesthetic felt-engagement, with its associated somatosensory anticipations, is mutually co-determining with our motor-sensory evaluations of action possibilities. In all our action, whether it be taking a step forward, reaching out tentatively with a hand, or gazing out over a landscape, we are continually, as part of our experiential horizon, asking tacit, pre-reflective, pre-noetic, non-propositionalized questions about our world and our being with and within it [9, 10]. Thus the feeling

[18] Clark provides the starkest example of an individual-centered cognitive approach in his Hypothesis of Organism-Centered Cognition (HOC): Human cognitive processing (sometimes) literally extends into the environment surrounding the organism. But the organism (and within the organism, the brain/CNS) remains the core and currently the most active element. Cognition is organism centered even when it is not organism bound. [7, p.139]

[19] Possibly a phrase originating in Plato's Phaedrus 265d-266a.

[20] For a thorough and engaging discussion of values, affordances, and value fields, see [26, 27, 46].

[21] The ambiguity with the term 'grasp' is intentional.

of being is, by its nature, a feeling of being *with*, the capacity for enkinaesthetic dialogue is an *a priori* nomological condition for agency, and, through the creation of kinaesthetic memories, melodies and imagination [48, 50], the generation of a felt anticipatory dynamics, makes possible the effective engagement with object- and movement-dependent sensorimotor contingencies [36]. In our intersubjective openness we don't just possess a transcendental intersubjectivity [58], we possess a transcendental enkinaesthesia.

5 ENKINAESTHESIA

The enkinaesthetic dialogue is rarely, if ever, simply two, though with the influence that language has had on our thinking we do tend to characterize it in this way. We exist within an ongoing processual dialogue from our earliest moments *in utero* to the time in which we cease to feel, and at that point others don't cease to feel, that is, to be enkinaesthetically linked to us. This is part of a universal dialogue that consists of a topologically complex web of relations of the community and reciprocity of sensing and experiencing agents and things and their felt implicit, and, sometimes, explicit intentional co-agency. It is this which co-constitutes conscious relations and the experientially recursive temporal dynamics of the non-symbolic, non-representationally-based experiential horizon for all agents.

> The organism does not develop in isolation from what happens around it; it is literally created (hence *poien*) by nature, while at the same time modifying both nature and itself. In this respect, autopoiesis more accurately describes what in the phenomenological structure of Paarung is generally presented as an experiential circularity, because the former stresses that the autonomy of the living [being] is the very result of its contextual dependence. [17, p.179-80]

Enkinaesthesia may emphasise the neuromuscular dynamics of the agent, the givenness[22] [24] of its experience, but it also emphasises the entwined, blended and situated co-affective phenomenological structure of *Paarung*. Unlike the circularity that characterizes *Paarung* enkinaesthetic activity possesses a recursive dynamics, and it is these experientially recursive temporal dynamics that lead to the formation and maintenance of integral enkinaesthetic structures and melodies. Such deeply felt enkinaesthetic melodies emphasise the dialogical nature of the feeling of being as necessarily having the feeling of being-*with*, being-*among*, or even being-*in-with*, and demonstrate the paucity of those notions that individuate agents and objects and treat them as singular and independent, as states and substances.

> If one wants to speak of a commitment to the alive consciousness of others here, one should speak not of a cognitive commitment but, rather, of a practical commitment. Like the baby in relation to her mother, we are involved with each other. It is our joint cohabitation that secures our living consciousness for each other. We live and work together. [37, p.33]

It is certainly our 'cohabitation', our being in affective relations of community and reciprocity, that secures our living consciousness for one another; the pragmatics of the commitment, of the living and working together are, in a strong sense, to do with survival. But "our living consciousness for one another" is just one element of a much broader 'practical commitment' expressed throughout the enkinaesthetically co-ordinating, values-realising ongoing processual situation which comprises no well-defined boundaries between agents, actions, substances, and objects. It is a 'practical commitment' which emphasises the bodily, kinaesthetic affective tonalities that underpin and make possible the proto-modal in relationships, or what Gendlin calls the "implicit interactional bodily intricacy".

> There is an implicit interactional bodily intricacy that is first – and still with us now. It is not the body of perception that is elaborated by language, rather it is the body of interactional living in its environment. Language elaborates how the body implies its situation and its next behaviour. We sense our bodies not as elaborated perceptions but as the body sense of our situations, the interactional whole-body by which we orient and know what we are doing. [22, p.352]

What seems to be missing from both authors, Noë and Gendlin, is the reciprocal co-affectivity of these feeling states in the co-creation of the interactional dialogue. Such co-affectivity is characterized by being inherently intentional, which is to say that being-*with* and being-*among* is necessarily relational and comes already clothed in 'aboutness', already saturated with intentionality.[23] The 'knowing', referred to by Gendlin, occurs through the enkinaesthetic affective enfolding which enables the balance and counter-balance, the attunement and co-ordination of whole-body action through mutual, let's say, reciprocal adaptation. It is this that Maturana refers to as 'languaging'.

> To language is to interact structurally. Language takes place in the domain of relations between organisms in the recursion of consensual coordinations of actions, but at the same time language takes place through structural interactions in the domain of the bodyhoods of the languaging organisms. . . . As the body changes, languaging changes; and as languaging changes the body changes. [34, §9.5]

Thus we exist in a continuous flow of the creation and fragmentation of agential-kinaesthetic, interpersonal, intersubjective, intercorporeal, enkinaesthetic melodies. But we must be clear that enkinaesthesia is not simply empathy by another name or in another guise.

Whilst I agree wholeheartedly with the claim that "Individual human consciousness is formed in the dynamic interrelation of self and other, and therefore is inherently intersubjective." [52, p.1], it is neither just at the level of persons or selves that this interrelation occurs, and nor is is restricted to human consciousness [15]. Additionally, and more importantly, there is a strong sense in which the individual at a sub-personal level is never an individual; at this level of explanation the agent is a co-constituent of an enkinaesthetically co-ordinating, values-realising situation, and it is the situation, with its pico-scale affective and motor resonances, which needs to become the focus of our attention. Current work in empathy, even at the

[22] We might understand self-givenness in terms of Husserl's concept of "eidetic intuition": the direct givenness which "refers to the acts in which 'objects show up in person'" [18, p.45] and which primarily reveals itself as a perceptual and imaginative act concerned with disclosing an essence [ibid., p.55]. Self-givenness is concerned with the revelation of the tight experiential coupling between body and ownership of the experience.

[23] Husserl speaks of the *hyle* or *hyletic* as the sensuously palpable affection in a temporal horizon of subjective bodily living [31, §85, pp.203-7], but he seems inconsistent in terms of the relational dialogue for he denies the intentionality of the sensuous saying: "the *sensuous, which has in itself nothing pertaining to intentionality*" [ibid. p.203]. But if the hyletic core affection is non-intentional, as Husserl would seem to think, we would be unable to establish kinaesthetic memories, melodies, and anticipations but, much more seriously we would be ineffective socially, unable to engage enkinaesthetically through the "passive synthesis" of affective enfolding.

lowest level of its conception as spontaneous and unreflective motor mimicry [15, p.39], fails to address this in sufficient detail; it also fails to prioritise the situation over the individual when it is no longer the individual which should be the primary concern.

Co-dependent enkinaesthetic affectivity must be primitive to and necessary for empathy. As Cowley says "First, we interact and co-engage; later, as persons, we construe experience." [11]. Empathy is felt at an intersubjective personal level and is one of the ways in which we construe experience, but that construal emerges through a topologically complex dialogical array of affective co-agency, that is, the rich enkinaesthetic co-engagement of perpetual situations. The intentional arc of action is not the means to put ourselves into *a* situation [35] but the means with which we propel ourselves fallenly and thrownly, in the Heideggerian senses, through *the* – experientially entangled, continuously folding, enfolding and unfolding – situation which constitutes our *Lebenswelt* with its inevitable *Mitseinwelt*.

6 ENKINAESTHESIA AND THE ETHIOSPHERE

The 'situations' agents inhabit possess, what Steinbock [47] refers to as, affectively "saturated intentionality". It is through the intentionally saturated affectively-laden enkinaesthetic engagement that things and others in our *Mitseinwelt* are felt as concerns for us. We reach, touch, taste, grasp, hear, and see, and all as the felt mattering of spontaneously occurring motor and aesthetic evaluation. We may speak of things and agents but it is at the level of perceptual and kinaesthetic experience that we are primordially related to our world. It is at the level of textures, smells, tastes, colours, movement, and so on, that we check out our world, asking non-cognitive, pre-reflective questions about whether it will continue as it feels now, anticipating how it might change and how it would feel if it does, and being most keenly aware of ourselves, not when it all runs smoothly but, when our anticipations are confounded. All of this rich experiential tapestry is woven through with the primordial moods of care and the openness to the possibility of fear. Through our ongoing processual enkinaesthetic dialogue we project ourselves into our possibilities; we grasp – with our hands, our eyes, and our heads – the perceptual-kinaesthetic values and facticity of our "Being- already-in-(the world) as Being-alongside (entities encountered within the world)" [23, p.237].

Thus, it is within the continuous flow of the creation and fragmentation of agential-kinaesthetic, interpersonal, intersubjective, intercorporeal, enkinaesthetic melodies that these situations constitute "our living consciousness for one another"; and it is the felt reciprocity of active forces between agents – again, the enkinaesthetic – which "over-determines" the relational community and which expresses, in their embodied affective manner, the vacillations of freedom, commitment, and responsibility felt in our temporally recursive relationships with one another. We are endogenously ethical, folding enkinaesthetically into the being-in-time of the other.

So, if agents are, as I have argued, transcendentally enkinaesthetic, and our actions *ab initio in utero* are felt concernful matterings which are values-realising, then the domain of values, what I shall here call the 'ethiosphere', is co-extensive with the domain of that which we deem to matter and have meaning for us, that is, it is co-extensive with the semiosphere. Since "Every action ... that consists of perception and operation imprints its meaning on the meaningless object" [55, p.31], and every action is affectively-laden co-engagement replete with concernful, values-realising exploration, every action, even *in utero*, operates within the semiosphere, the immanent habitus, and that semiosphere extends throughout the biosphere.

> The semiosphere is a sphere like the atmosphere, the hydrosphere or the biosphere. It penetrates these spheres and consists in communication: sounds, odours, movements, colours, electric fields, waves of any kind, chemical signals, touch etc. [28, p.35]

Our senses open us to the reception of these forms of communication, but it is not a passive reception; it is a reciprocally affective, intentional, co-agential, concernful, enkinaesthetic communication in which we are able to affect others and be affected by them, to move and be moved within the sphere of ethical engagement, that is, within the ethiosphere. The implications for the community and reciprocity of enkinaesthesia and its ranging over the ethiosphere, all that is sensed and felt by the agent, is significant. Although the details of this significance cannot be drawn out at length here, there are a number of instances which can be presented as a means to advance the claim.

Merleau-Ponty, though ostensibly speaking about the conscious 'cognitive' relation whilst I have emphasised the dialogical 'felt' one, recognises the fragmentation and failure of the intentional enkinaesthetic relation in illness.

> Let us therefore say ... that the life of consciousness-cognitive life, the life of desire or perceptual life-is subtended by an "intentional arc" which projects around about us our past, our future, our human setting, our physical, ideological and moral situation, or rather which results in our being situated in all these respects. It is this intentional arc which brings about the unity of the senses, of intelligence, of sensibility and motility. And it is this which 'goes limp' in illness. [35, p.136]

In all organisms the intentional agency directed towards another, that is, its conscious affective, enkinaesthetic relation to the other, is seeking to affect the other and be affected by the other. As Ratcliffe states "practical *relatedness* between people [is] an aspect of interpersonal understanding and experience that typifies most social encounters" [42, p.196] and through which "patterns of affective interaction between people" [ibid. p.197] are established which make mutual understanding possible. So, if the organism, whilst functionally fit, attempts to fold into its environment but the environment fails to perceive it or perceives it but fails to respond to it – say in the case of being rendered socially invisible through ostracism – the organism will feel the failure in affective response; the enkinaesthetic entwining, blending and situating co-affective feeling of the presence of the other will be absent, and the negation of affect will be felt as suffering. This is a case of social ailment that ruptures the enkinaesthetic relatedness and produces a real bodily affective disorder, but Ratcliffe presents a cogent account of how changes in "existential feeling, involving the diminution or absence of possibilities for interpersonal relatedness" [2008a p.143] can be presented as explanations for Cotard's and Capgras' syndromes and depersonalisation. [24]

[24] Cotard's delusion was first identified by Jules Cotard in 1880 and is characterised as *délire des négations*, that is, the delusion that one is dead or that the world no longer exists. [See, for example, [57, 1, 41].] Capgras delusion is characterised by a person's being able to recognise a family member or friend as a family member or friend, but simultaneously believing the person to be an impostor. [See, for example, [4, 39, 25, 41, 42]. It is now proposed that in both instances there is a malfunction between the face recognition areas of the brain (the fusiform gyrus) and areas associated with emotional recognition, for example, the amygdala and other limbic structures. The face may be recognised by the proper functioning fusiform structures but, because of the faulty connection, it lacks the usual affective accompaniment that generates emotional recognition.

Significantly for the robustness of the claim for an enkinaesthetic dialogue, Ratcliffe argues that none of these illnesses results from affective diminution alone, rather it is an affective diminution that results in the fragmentation of the practical relatedness and loss of possibilities for attunement. The person who suffers from Cotard, Capgras or depersonalisation is, to varying degrees, incapable of feeling the reciprocally affective, intentional, co-agential, concernful existential feelings of Being-alongside. The diminution of affect dims the capacity for motor-aesthetic value-realising engagement, reducing their field of concernful mattering, and damaging their well-being and overall functional fitness. As their ethiosphere shrinks, so shrinks their field of engagement, their semiosphere.

In the context of non-communicative states we find another excellent example of the potential explanatory power of enkinaesthesia.

The sensation or feeling we have of Being-alongside, the *Mitseinwelt*, is conscious within topologically complex affectively-laden dialogical fields. In some dream states we are able to create these fields in the absence of actual waking experience; REM sleep and Lucid Dreaming offer such possibilities, but in a coma this ability seems lost. In a coma or under general anaesthesia there is a temporary cessation of the normal practical relatedness we feel; our level of arousal and our awareness of the environment and ourselves is low to non-existent [38]. There is an absence of existential feeling. However, in, for example, a minimally conscious state[25] which possesses a higher level of arousal and, in some cases, a greater level of awareness of the environment and the self, some relatedness continues to be present and it might be possible to create other forms of relatedness by suggestion from outside [33]. The problem in these cases is how to disentangle the automatic brain activation from the conscious intentionally-related activation.

> Owen *et al*. (2006) have recently addressed this issue by asking non-communicative patients to actively perform mental imagery tasks. In one exceptional VS patient studied five months after a cerebral trauma, activation was observed in the supplementary motor area after being asked to imagine playing tennis. When asked to imagine visiting the rooms of her house, activation was seen in premotor cortex, parahippocampal gyrus and posterior parietal cortex. Near identical activation was observed in the 34 healthy volunteers studied in Cambridge and Liège. The patients decision to imagine playing tennis rather than simply rest must here be seen as an act of willed intention and, therefore, clear evidence for awareness. [33, p.735]

Whether or not this patient did 'decide' to play tennis is not important in this paper, what is important is the relatedness which was becoming possible. It is with the possibility of relatedness that the world reopens to us, disclosing itself once more as the arena for enkinaesthetic dialogue, concernful exploration and a field of values-realising possibilities. This corresponds well with what Laureys *et al.* [33] say next: "Interestingly, when re-examined six months later the patient showed inconsistent visual tracking–the most frequently encountered clinical sign of recovery from VS". [ibid.] This patient's recovery coincides with an improvement in their enkinaesthetic awareness and, thus, their motor-aesthetic values-realising engagement.

To the case of illness we can also add the loss of the reciprocal felt intentional relation in grieving for the death of another. For example, a friend has recently lost her cat to cancer, she grieves for the loss. Perhaps we might explain the grief in the following way: the dialogical relation she had with Sara [the cat] continues even though Sara has gone. She still thinks about Sara, expects to see her and to reach out to touch her, and anticipates her purr and her vocalisations. In part my friend's grief is a result of the absence of the habituated enkinaesthetic feedback in that topologically rich dialogue. We become used to interacting with and being affected by the other, anticipating and receiving feedback from the other, like the purring that the cat does when we stroke it, but that deeply felt reciprocated response is no longer there to be received. The co-engagement is absent, and the lack of affective feedback is felt quite simply as negative affect and loss.

One brief last word is important to respond to a possible objection and emphasise the range over which the ethiosphere can be said to extend. Tønnessen [54] distinguishes between the semiotic niche or semiosphere, and an ontological niche. The semiotic niche, he argues, operates within the class of ideal agents, and the ontological niche describes real agential relations. So, the ontological niche concerns living organisms. If we accept his distinction, then the ethiosphere would seem most naturally to apply in real world circumstances where the relations are felt concernful matterings and not over the semiosohere; however, from an enactivist ethical consideration of real, multi-directional, contrapuntal relations [8], it would be possible to conceive of, and even formulate, a normativity that ought to hold in ideal circumstances and, thus, across the semiosphere. So, although at first glance the notion of the ethiosphere seems more clearly co-extensive with the non-ideal ontological niche, there is no confounding reason to think it not, at least, potentially co-extensive with the semiosphere as well.

7 CONCLUSION

I have argued that the capacity for enkinaesthetic dialogue is an *a priori* nomological condition for agency and the generation of a felt anticipatory dynamics both within and between agents. It is not empathy but it is a necessary requirement for empathy. It corresponds in some ways to the 'existential feeling' spoken of by Ratcliffe [41, 42] but the emphasis in enkinaesthesia is on the dialogue: the topologically complex web of relations of the community and reciprocity of sensing and experiencing agents and things and their felt implicit, and, sometimes, explicit intentional co-agency. Enkinaesthetic dialogical-relations are the preconceptual, prenoetic, experientially recursive temporal dynamics which form the deep extended melodies of relationships-in-time, and any understanding of how those relationships work, when they falter, when they resonate sweetly, and so on, will depend on a grasp, not only of our intersubjectivity or our intercorporeality but, of our enkinaesthesia. In arguing for this I hope to have demonstrated how these enkinaesthetic melodies emphasise the dialogical nature of the feeling of being as the feeling of being-*with* or being-*among*, and to have demonstrated the paucity of individuating notions that treat agents as isolated and singular. I hope also to have shown the explanatory power and potential that enkinaesthesia has in the contexts of "mental" illness, health and caring.

[25] Minimally conscious states can include some kinds of vegetative state, more normally those described as 'persistent' not 'permanent'. [33]

REFERENCES

[1] G E Berrios and R Luque, 'Cotard's syndrome: Analysis of 100 cases', *Acta Psychiatrica Scandinavia*, **91**, 181–188, (2995).

[2] A. M. Borghi and F. Cimatti, 'Embodied cognition and beyond: Acting and sensing the body', *Neuropsychologia*, **48**(3), 763–73, (2010).

[3] *On Being Moved: From Mirror Neurons to Empathy*, ed., S Bråten, John Benjamins, 2007.

[4] J Capgras and J Reboul-Lachaux, 'Illusions des 'sosies' dans un délire systématisé chronique', *Bulletin de la Société Clinique de Médicine Mentale*, **2**, 6–16, (1923).

[5] D. B. Chamberlain. Communication before language [online]. 1995. URL: http://www.birthpsychology.com/lifebefore/comm.html.

[6] D. B. Chamberlain. The fetal senses: A classical view [online]. 1997. URL: http://www.birthpsychology.com/lifebefore/fetalsense.html.

[7] A. Clark, *Supersizing the Mind: Embodiment, Action, and Cognitive Extension*, Oxford University Press, Oxford, 2008.

[8] G Colombetti and S Torrance, 'Emotion and ethics: and inter-(en)active approach', *Phenomenology and the Cognitive Sciences*, **8**(4), 505–26, (2009).

[9] Rodney M. J. Cotterill, 'On the unity of conscious experience', *Journal of Consciousness Studies*, **2**(4-5), 290–311, (1995).

[10] Rodney M J Cotterill, *Enchanted Looms: Conscious Networks in Brains and Computers*, Cambridge University Press, 1998.

[11] Stephen J Cowley. The life and mind seminar network [online]. 2008. URL: http://lifeandmind.wordpress.com/battle08/battle08-participants/ [cited March 2013].

[12] Antonio R. Damasio, *Descartes' Error: Emotion, Reason and the Human Brain*, Picador, 1994.

[13] Antonio R. Damasio, *The feeling of what happens: Body, emotion and the making of consciousness*, Harcourt Brace, New York, 1999.

[14] Antonio R. Damasio, *Looking for Spinoza: Joy, sorrow, and the feeling brain*, Harcourt Brace, New York, 2003.

[15] F de Waal and E Thompson, 'Primates, monks and the mind', *Journal of Consciousness Studies*, **12**(7), 38–54, (2005).

[16] Gilles Deleuze and Félix Guattari, *A Thousand Plateaus: Capitalism and Schizophrenia*, University of Minnesota Press, Minneapolis, 1987.

[17] Natalie Depraz, 'The rainbow of emotions: at the crossroads of neurobiology and phenomenology', *Continental Philosophy Review*, **41**(2), 237–59, (2008).

[18] Natalie Depraz, F Varela, and P Vermersch, *On Becoming Aware: Steps to a Phenomenological Pragmatics*, Advances in Consciousness Research Series, John Benjamins, Amsterdam, 2003.

[19] Chris Dobbyn and Susan A J Stuart, 'The self as an embedded agent', *Minds and Machines*, **13**(2), 187–201, (2003).

[20] Gerald Edelman, *Bright Air, Brilliant Fire: On the Matter of the Mind*, Basic Books, New York, 1992.

[21] Gerald Edelman, *Second Nature: Brain Science and Human Knowledge*, Yale University Press, 2006.

[22] Eugene T Gendlin, 'The primacy of the body, not the primacy of perception', *Man and World*, **25**(3-4), 341–53, (1992).

[23] Martin Heidegger, *Being and Time*, SCM Press, London, 1962.

[24] Michel Henry, *L'Essence de la manifestation/The Essence of Manifestation*, Nijhoff, The Hague, 1963.

[25] W Hirstein and V S Ramachandran, 'Capgras syndrome: a novel probe for understanding the neural representation of the identity and familiarity of persons', *Proceedings of the Royal Society of London, B: Biological Sciences*, **264**, 437–44, (1997).

[26] Bert H Hodges, 'Values define fields: the intentional dynamics of driving, carrying, leading, negotiating and conversing', *Ecological Psychology*, **19**(2), 153–78, (2007).

[27] Bert H Hodges and R M Baron, 'Values as constraints on affordances: Perceiving and acting properly', *Journal for the Theory of Social Behaviour*, **22**(3), 263–94, (2007).

[28] Jesper Hoffmeyer, 'The global semiosphere', in *Semiotics Around the World: Proceedings of the 5th IASS Congress, Berkeley*, eds., Irmengard Rauch and Gerald F Carr, pp. 933–936, Berlin/New York, (1995). Mouton de Gruyter.

[29] Jesper Hoffmeyer, *Biosemiotics, An examination into the Signs of Life and the Life of Signs*, University of Scranton Press, Scranton and London, 2008. trans. J. Hoffmeyer and D. Favareau.

[30] Edmund Husserl, *Analysen zur passiven Synthesis: Aus Vorlesungs- und Forchungsmanuskripten 1918–1926*, Martinus Nijhoff, The Hague, 1966.

[31] Edmund Husserl, *Ideas Pertaining to a Pure Phenomenology and to a Phenomenological Philosophy, First Book. General Introduction to a Pure Phenomenology*, Springer, 1982. trans. Fred Kersten.

[32] Immanuel Kant, *The Critique of Pure Reason*, Macmillan Press, 1929.

[33] S Laureys, F Perrin, and S Brèdart, 'Self-consciousness in non-communicative patients', *Consciousness and Cognition*, **16**, 722–41, (2007).

[34] Humberto R. Maturana, 'Ontology of observing: the biological foundations of self-consciousness and the physical domain of existence', in *Conference Workbook for Texts in Cybernetic Theory*, ed., R. E. Donaldson, American Society for Cybernetics, (1988).

[35] Maurice Merleau-Ponty, *Phenomenology of Perception*, Routledge & Kegan Paul, London, 1962.

[36] Alva Noë, *Action in Perception*, MIT Press, 2004.

[37] Alva Noë, *Out of Our Heads, Why You Are Not Your Brain, and Other Lessons from the Biology of Consciousness*, Hill and Wang, New York, 2009.

[38] F Plum and J B Posner, *The Diagnosis of Stupor and Coma*, Wiley, 3rd edn., 1983.

[39] V S Ramachandran and S Blakeslee, *Phantoms in the Brain: Probing the Mysteries of the Human Mind*, William Morrow, New York, 1998.

[40] Matthew Ratcliffe, 'Heidegger's attunement and the neurophysiology of emotion', *Phenomenology and the Cognitive Sciences*, **1**(3), 287–312, (2002).

[41] Matthew Ratcliffe, *Feelings of Being: Phenomenology, Psychiatry and the Sense of Reality*, Oxford University Press, Oxford, 2008.

[42] Matthew Ratcliffe, 'The phenomenological role of affect in the Capgras delusion', *Continental Philosophy Review*, **41**, 195–216, (2008).

[43] Maxine Sheets-Johnstone, *The Primacy of Movement*, John Benjamins, Amsterdam/Philadelphia, 1999.

[44] Maxine Sheets-Johnstone, 'Kinetic tactile-kinesthetic bodies: Ontogenetical foundations of appreticeship learning', *Human Studies*, **23**, 243–70, (2000).

[45] Maxine Sheets-Johnstone, 'Kinesthetic memory', *Theoria et Historia Scientiarum*, **7**, 69–92, (2003).

[46] Sune V Steffensen and Bert H Hodges, 'The ecology of values-realizing in dialogical and social systems', in *Expression, Engagement Embodiment: The Ecology of Situation Transcendence, University of Glasgow*, (February 2010).

[47] Anthony J. Steinbock, *The Body: Classic and Contemporary Readings*, chapter Saturated Intentionality, 178–99, Blackwell, 1999.

[48] Susan A. J. Stuart, 'Machine consciousness: Cognitive and kinaesthetic imagination', *Journal of Consciousness Studies*, **14**(7), 141–53, (2007).

[49] Susan A. J. Stuart, 'The mindsized mashup mind isn't supersized after all', *Analysis*, **70**(1), 174–83, (2009).

[50] Susan A. J. Stuart, 'Conscious machines: Memory, melody and imagination', *Phenomenology and the Cognitive Sciences*, **9**(1), 37–51, (2010).

[51] Susan A J Stuart and Paul J Thibault. Enkinaesthestic polyphony as the underpinning for first-order languaging. Forthcoming, 2013.

[52] Evan Thompson, 'Empathy and consciousness', *Journal of Consciousness Studies*, **8**(5-7), 1–32, (2001).

[53] Evan Thompson, *Mind in Life: Biology, Phenomenology and the Sciences of Mind*, Harvard University Press, 2007.

[54] M Tønnessen, 'On contrapuntuality: Semiotic niche vs. ontological niche, the case of the scandinavian wolf population', in *Contrapuntuality – Gatherings in Biosemiotics, Prague*, (2009).

[55] Jakob von Uexküll, 'The theory of meaning', *Semiotica*, **42**(1), 25–82, (1982). doi:10.1515/semi.1982.42.1.25.

[56] Rachel Wood and Susan A. J. Stuart, 'Aplasic phantoms and the mirror neuron system: an enactive, developmental perspective', *Phenomenology and the Cognitive Sciences*, **8**(4), 487–504, (2009).

[57] A W Young, K M Leafhead, and T K Szulecka, 'The Capgras and Cotard delusions', *Psychopathology*, **27**, 226–31, (1994).

[58] Dan Zahavi, 'Horizontal intentionality and transcendental intersubjectivity', *Tijdschift voor Filosofie*, **59**(2), 304–321, (1997).

Situating enactive processes or placing the observer back in the scene: a case for the empirical study of perception

Etienne B. Roesch [1] and **Carol MacGillivray** [2] and **Bruno Mathez** [2] and **Frédéric Fol Leymarie** [2]

Abstract. Most of what is known about visual perception in humans is due to empirical work relying on computerised display. In this position paper, we argue that this technology does not permit the investigation of the enactive processes that support visual experience. We discuss the requirements for investigation of situated, embodied interaction of an observer with an observed scene, and introduce the Diasynchronoscope, an art-form technique that permits the exploration of perception *in situ*. For the enactive paradigm to be successful, we argue, researchers need to be aware of the biases introduced by the technology used in empirical work, and explore alternative ways to probe enactive cognition.

1 INTRODUCTION

In its most notorious formulation, enaction bridges experience and the outside world by emphasising the role played by the body in the meaningful interaction of the organism with the environment [18]. As such, this extension to cognitive science contrasts with the more mainstream paradigm that focuses almost exclusively on the brain, and draws from concepts of the computational theory of mind, like information processing *modules* and *representations* as internal models of the outside world [3, 4, 16]. In this position paper, we argue that current empirical methods to the study of visual perception reinforce the computational paradigm, and do not permit the investigation of the enactive processes that give rise to visual experience. We review hardware, software and contextual considerations that may impact on what is available to researchers in a typical empirical setup, and then present the Diasynchronoscope, an animation technique that makes the investigation of perception *in situ* possible [8, 9].

Enactive processes pertain to the closed-loop operations that ground the biological organism in the environment, progressively supporting the most optimal way to perceive and react to stimulations [15]. In the case of visual perception, these processes encompass the interaction between reflections of light and the perceiving organism, as well as the sense of embodiment and situatedness experienced by the organism, brought forth by a process of sensorimotor exploration [13], homeostasis (in the context of autopoiesis, see [11]), proprioception and the indelible experience of the physics of the world. Through this slow and maturing process, the biological substrate adjusts itself to the demands of the environment, to reach a state of balance that will ground and support the interaction of the organism with the environment. These adjustments occur at many levels, including the central [2, 1] and the peripheral nervous systems. Visual perception then results from the interaction between the many levels of processing in the nervous system.

Models of visual perception often describe the different stages of processing that a visual percept must go through, often sequentially, to reach awareness and be acted upon by an observer; see for instance [14]. This theoretical work originates from empirical studies that place the observer in highly controlled situations, with the hope of disentangling these so-called computational stages of processing [10]. These controlled situations often approximate the interaction with the environment an observer is believed to experience, with the help of technology. Typically, a scene will be simulated and presented on a display, i.e. a computer monitor, and the observer is asked to perform some task, e.g. a discrimination task. It is understood, but often forgotten, that this simulation process only recreates an incomplete representation of the richness that composes the interaction of the observer with the perceived scene. To study stereopsis, for instance, which is the ability to experience 3D vision, experimenters will derive a 2D representation of a 3D scene [17], to be displayed on a 2D monitor or at best on a 3D virtual reality display, thereby losing a range of information related to the physical presence of the observer inside the observed scene, e.g. the experience of gravity, the intrinsic balance of the body or the physics underlying light propagation that forms natural colors.

Even though using these methods have been very informative, yielding a detailed description of some of the processing stages of visual information, we argue that a successful approach to investigating perception will not limit the field of study to simulated situations. Instead, greater emphasis should be placed on the construction of empirical situations that resembles natural interaction as closely as possible, while ensuring the quality criteria intrinsic to the scientific method, e.g. reproducibility, precision, accuracy and hypothesis testing. In the following, we present the Diasynchronoscope, an animation technique that has been used as an art form to recreate Gestalt-based perceptual phenomena, and which, we argue, may allow researchers to probe perception *in situ*.

2 THE DIASYNCHRONOSCOPE TECHNIQUE

The name Diasynchronoscope[1] combines "diachronic", the study of a phenomenon as it evolves over time, with "synchronous" and "scope" (view), to evoke the experience that this technique generates for the observer [8, 9]. The Diasynchronoscope employs a simple

[1] University of Reading, UK
– Cybernetics, School of Systems Engineering
– Centre for Integrative Neuroscience and Neurodynamics
Correspondence: contact@etienneroes.ch
[2] Department of Computing, Goldsmiths, University of London, UK

[1] http://www.doc.gold.ac.uk/diasynchronoscope

technique to create animation out of inanimate objects. Each sculptural installation is composed of objects arranged in space so as to represent the unfolding of an animation over time. The objects are shaped to represent the interaction of the object with the environment (e.g., using the animation tropes of squash and stretch to represent weighted interaction with the environment, and enhance the impression of speed). The sculptures thus consist of multiple objects representing a single object. The objects are then lit selectively and sequentially, unveiling one object at a time, thereby creating the embodied equivalent to an animated cartoon (see Figure 1). The observer is placed inside the scene, and perceives a continuity in the interaction of the object with the environment, as if observing a moving object evolving in space and time. The perception can be reproduced very precisely and as often as necessary.

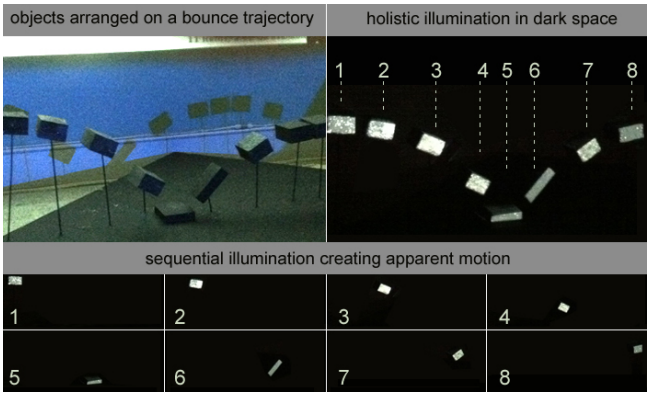

Figure 1. Design of a simple Diasynchronoscope animation representing a bouncing cube. The figure shows the whole installation. Notice the distortion of the shape of the cube, and the spacing between each of the cubical blocks that give the impression of speed, rigidity/flexibility and interactivity with the environment; here, the ground plane of the scene. The closer the blocks are together, the slower the impression of movement. During the animation, each object is lit sequentially, leading the observer to perceive the continuous interaction of the cube with the environment. The animation runs at 24 frames per second (fps) where each frame represents 41.6 milliseconds; the first four blocks and the last two blocks are illuminated for 83.2 ms (2 fps), block five (the squashed block) is lit for 124.8 ms (3 fps), and the sixth block for 20.8 ms (1 fps).

We contrast this novel technique against visual displays of the type most often used in perceptual experiments in psychophysics, psychology and neuroscience, namely computer monitors. Importantly, transposing a natural scene to such a display implies complying with constraints at both hardware and software levels, which may have a significant impact on the quality of the resulting 2D representation.

First, these devices vary in (viewable) screen size, luminance, display resolution, refresh rate (temporal resolution), contrast ratio, viewing angle and ICC[2] colour profile, amongst other criteria. Colour display is achieved through the approximation of the distribution of light power against wavelengths, through electro-chemical mechanisms: e.g., by combining red, green and blue phosphors, excited by an electron beam, in the case of traditional cathode ray tube monitors (CRTs).

Second, in a typical perceptual experiment, the participant (observer) is seated at a distance of about 60 cm away from the monitor, presented with a display of about 30 visual angle, in a dimly lit room.

[2] International Color Consortium (ICC), an international consortium responsible for the definition and diffusion of open, vendor-neutral and cross-platform standards for colour management systems.

The experimenter proceeds with displaying stimuli of interest on the monitor. These visual stimuli result from the calculation of the best 2D representation possible of the 3D scene to be displayed, which is then rendered on the display. This calculation takes into account optical conditions, such as stereopsis, parallax, perspective or field of view, computed against idealised laws of physics if needed.

The nature of this setup permits the fine control of empirical situations, by providing researchers with ways to control for the effects of the parameters that are believed to play a role in perception. The ability to manipulate the material recreated, however, comes with a significant overhead related to the necessity to comply with the constraints imposed by the technology, both at the hardware and the software levels. For instance, on CRTs, images displayed go through a rendering process constrained by the refresh rate of the raster beam; this process creates flickering, which creates perceptual noise. Even if this flickering may not be consciously reported by observers, evidence shows that it is perceived by the perceptual system and will thus impact perception [7]. Newer displays, like modern LCDs or Thin Film Transistor technology, do not exhibit flickering due to raster beam, but suffer extremely poor accuracy and reliability that make them unusable for perceptual research [19].

Taking these elements into account, arguably the perception experienced by the observer in such a situation is somewhat remote from the more natural one that sees the observer situated *inside* the observed scene. In contrast, techniques like the Diasynchronoscope quite literally take perception out of the box, by providing researchers with a screen-less method to empirically study perception. Our preliminary work shows that it is possible to create replicable, screen-less animation in such a way that the observer perceives a continuous flow of perception, supporting a communication with their environment that closely resembles the interaction with a real-life, at scale moving object.

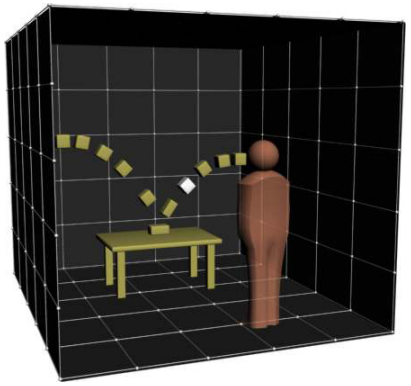

Figure 2. Schematic representation of an installation within the Diasynchronoscope technique–here a bouncing cube. The observer is standing up in the middle of a dark room, and perceives the animation as the objects are lit sequentially around them. By standing up in the middle of the observed scene, the experience of the observer is enriched by their perception of gravity, the intrinsic motion due to their balance, and their positioning in the scene. In situation, only one object (in white in the picture) is illuminated at a time; the room is completely dark and the other objects (in orange in the picture) are not visible.

Furthermore, a typical installation created within the Diasynchronoscope technique will see the observer standing up in the middle of the scene to be observed. In one installation, for instance, a bouncing ball may progress from one side of the room to the other, reacting to obstacles on the way. By standing up in the middle of this

scene, the observer's visual experience will be enriched with their experience of their own interaction within the scene, as brought forth by gravity, for instance, their sense of balance and proprioception, which will continuously impact on their perception of the moving ball (see Figure 2). A growing body of results indeed shows that these aspects of embodiment may play a significant role in perception: The experience of gravity, for instance, shapes our expectations of trajectories in space and time, explaining why the temporal aspect of a ball moving upward will be experienced less reliably, compared to that of a ball moving downward [12].

Interestingly, taken together, the results from this latter set of experiments show that accurate perception of motion duration is enhanced for objects moving according to gravity constraints, i.e., downward, even though this effect was not very large, 10-20% depending on the experimental condition [12]. The authors further discuss the role of the cues provided to the participants on the computer screen, which signified the scale of the moving objects, and hypothesise that more realistic cues would yield much stronger effects [20]. In two of the six experiments conducted, the researchers tried to manipulate the experience of gravity itself by either tilting the seat of the participants or tilting the computer display by 45 degrees relative to Earth vertical. The authors replicated the trend of their results, but were unable to disentangle the role of experienced gravity in perception.

The Diasynchronoscope is by nature a situated and embodied mode of perception, and we thus posit that empirical work using this technique will permit to answer questions such as the role of experienced gravity in perception. By standing up in the middle of the scene, by being able to move freely and balance naturally, we hypothesise that the accompanying visual perception will be more accurate, more reliable and more precise, compared to a situation where the observer would be seated in front of a computer display. In addition to the effect of gravity standing up, versus seated on a chair, we hypothesise that the Diasynchronoscope will permit the investigation of aspects of perception related to stereopsis, parallax, depth, perceived time of motion and velocity, amongst others.

3 CONCLUSION AND DISCUSSION

In this position paper, we took a broad perspective over the empirical process that led to most of our knowledge about the visual system in humans. In particular, we critically analysed the role of computer displays in shaping the kinds of interactions that may be accessible to perceptual researchers. We argue that, in many ways, this technology hinders the exploration of the enactive processes at play in visual perception. We described hardware, software and contextual considerations that have a significant impact over the kind of information that is made available to researchers through empirical enquiry.

The enactive agenda in cognitive science emphasises the fundamental role of the perception–action loop for cognition, highlighting features of situatedness and embodiment. By providing researchers with the ability to animate objects in space and time, the Diasynchronoscope technique is ideally suited to layout the empirical conditions to investigate the effects of both of these aspects for cognition in general, and perception in particular [15]. Unlike the situations available to perceptual researchers through other technological means, like computer monitors, the observer is literally situated inside the observed scene, experiencing the full blown embodiment of their interaction with the environment. The observed scene is at scale, resembling as closely as possible life-size interactions that may occur with animated objects in space and time. Researchers are thus given the opportunity to observe and measure aspects of perception that are not readily available in typical empirical setups, like sensorimotor exploration in action or the role of proprioception in perception.

Other technologies may allow the creation of an immersive 3D environment, which could be used to probe some of the aspects of perception discussed in this paper, including 3D technology based on optical setups (e.g. anaglyphs/color filters or polarized lenses that separate the information available to both eyes), goggle-mounted virtual reality displays or cave automatic virtual environments (CAVE). Generally, these techniques are not easy to adapt to empirical research, requiring a lot of technical expertise and significant financial overheads. More importantly, none of these techniques allow for the creation of realistic interaction with objects in space and time in such a way that the observer *feels* embedded in a scene unfolding before their eyes[3].

Our efforts to develop new ways to probe enactive processes closely relates to work in the field of "experimental phenomenology", where researchers attempt to probe the lived aspects of the human mind. In particular, we propose that the Diasynchronoscope resonates with the concept of enactive interfaces proposed by Froese et al. [5] to describe devices that support "the enaction of new modes of experiences" [6]. Even though the Diasynchronoscope does not enter the original definition *sensu stricto*, which bounds the category to technological devices designed for augmented perception, we feel our goal is similar: going beyond the techniques and technologies currently available to researchers, and creating novel ways to explore the experiential aspects of the mind.

In conclusion, by pointing out aspects of perception that are not readily available to researchers with traditional empirical methods, which we think of as constitutive and core to the enactive process supporting perception, we assert that orthodox empirical methods bias our theoretical understanding of perception. Particularly, by imposing constraints on empirical designs, we propose that technology reinforces features of the computational paradigm in cognitive science, like stepwise, feed-forward information processing grounded in the passive reception of information by the mind; hence reinforcing static notions of domain-specific, encapsulated, automatic and ballistic information modules. Importantly, in this paper we described what might seem a conundrum for vision researcher, but we overtly suggest similar situations may arise in the investigation of other aspects of cognition.

ACKNOWLEDGEMENTS

We would like to thank Muriel Swijghuisen Reigersberg from Goldsmiths, University of London, for fruitful discussions and support, as well as the unwitting participants who experienced the first artistic/scientific experiments that helped refine the Diasynchronoscope into its current form.

REFERENCES

[1] E Di Paolo and H De Jaegher, 'The interactive brain hypothesis', *Frontiers in Human Neuroscience*, **6**, 1–16, (2012).
[2] Andreas K Engel, 'Directive minds: How dynamics shapes cognition', in *Enaction : toward a new paradigm for cognitive science*, 219–243, Cambridge, Mass. : MIT Press, (2010).
[3] Jerry A Fodor, *The Modularity of Mind*, An Essay on Faculty Psychology, The MIT Press, 1983.

[3] In fact, the closest device we imagine resembles the Diasynchronoscope technique might be *holodecks*, a fictional technology from the Star Trek universe located on starships and starbases, which can integrate physical and ethereal artefacts in simulation programs.

[4] Jerry A Fodor and Zenon W Pylyshin, 'Connectionism and Cognitive Architecture: A Critical Analysis', *Cognition*, **28**, 1–56, (1988).

[5] Tom Froese, Marek McGann, William Bigge, Adam Spiers, and Anil K Seth, 'The Enactive Torch: A New Tool for the Science of Perception', *IEEE Transactions on Haptics*, **5**(4), 365–375, (2012).

[6] Tom Froese, Keisuke Suzuki, Yuta Ogai, and Takashi Ikegami, 'Using Human–Computer Interfaces to Investigate 'Mind-As-It-Could-Be' from the First-Person Perspective', *Cognitive Computation*, **4**(3), 365–382, (2012).

[7] P Krolak Salmon, M A Hénaff, C Tallon Baudry, B Yvert, M Guénot, A Vighetto, F Mauguière, and O Bertrand, 'Human lateral geniculate nucleus and visual cortex respond to screen flicker', *Annals of neurology*, **53**(1), 73–80, (2003).

[8] C MacGillivray and B Mathez, 'Co-authored narrative experience: Affective, embodied interaction through combining the diachronic with the synchronistic', *Design and semantics of form and movement*, 62, (2012).

[9] C MacGillivray, B Mathez, and F Leymarie, 'Developing a system of screen-less animation for experiments in perception of movement', in *the Proceedings of the Eighth Annual Symposium on Computational Aesthetics in Graphics, Visualization, and Imaging*, pp. 71–80. Eurographics Association, (2012).

[10] David 1945-1980 Marr, *Vision : a computational investigation into the human representation and processing of visual information*, San Francisco : W.H. Freeman, 1982.

[11] Humberto Romesin Maturana and Francisco Varela, *Autopoiesis and Cognition: The Realization of the Living*, D. Reidel Publishing Company, Dordrecht, Holland, 1980.

[12] A Moscatelli and F Lacquaniti, 'The weight of time: Gravitational force enhances discrimination of visual motion duration', *Journal of Vision*, **11**(4), 5–5, (April 2011).

[13] J K O'Regan, *Why red doesn't sound like a bell: Understanding the feel of consciousness*, New York : Oxford University Press, 2011.

[14] Tomaso Poggio, 'The computational magic of the ventral stream', *Nature Precedings*, (2012).

[15] Etienne B Roesch, Slawomir J Nasuto, and Mark J Bishop, 'Emotion and anticipation in an enactive framework for cognition (response to Andy Clark)', *Frontiers in Psychology*, **3**, 1–2, (2012).

[16] Matthias Scheutz, *Computationalism*, New Directions, The MIT Press, 2002.

[17] Clifton M Schor, 'Stereopsis', in *The Visual Neuroscience: Volume 1*, 1300–1312, MIT Press, Cambridge, MA, (2004).

[18] Francisco J Varela, Evan Thompson, and Eleanor Rosch, *The embodied mind : cognitive science and human experience*, Cambridge, Mass. : MIT Press, 1991.

[19] Stefan Wiens, Peter Fransson, Thomas Dietrich, Peter Lohmann, Martin Ingvar, and Arne Öhman, 'Keeping it short: a comparison of methods for brief picture presentation.', *Psychological science*, **15**(4), 282–285, (April 2004).

[20] Myrka Zago, Gianfranco Bosco, Vincenzo Maffei, Marco Iosa, Yuri P Ivanenko, and Francesco Lacquaniti, 'Internal models of target motion: expected dynamics overrides measured kinematics in timing manual interceptions.', *Journal of neurophysiology*, **91**(4), 1620–1634, (April 2004).

20 Years After The Embodied Mind–Why is Cognitivism Alive and Kicking?

Vincent C. Müller [1]

Abstract. I want to suggest that the major influence of classical arguments for embodiment like "The Embodied Mind" by Varela, Thomson & Rosch (1991) has been a changing of positions rather than a refutation: Cognitivism has found ways to retreat and regroup at positions that have better fortification, especially when it concerns theses about artificial intelligence or artificial cognitive systems. For example: a) Agent-based cognitivism' that understands humans as taking in representations of the world, doing rule-based processing and then acting on them (sense-plan-act) is often limited to conscious decision processes; and b) Purely syntactic cognition is compatible with embodiment, or supplemented by embodiment (e.g. for 'grounding'). While the empirical thesis of embodied cognition ('embodied cognitive science') is true and the practical engineering thesis ('morphological computation', 'cheap design') is of- ten true, the conceptual thesis ('embodiment is necessary for cognition') is likely false - syntax is often enough for cognition, unless grounding is really necessary. I conclude that it has become more sensible to integrate embodiment with traditional approaches rather than "fight for embodiment" or "against cognitivism".

1 Cognitivism / Computationalism

The classic view of what is called 'cognitivism' or, more accurately, 'computationalism' is that syntactic processing over symbolic representation is sufficient for intelligence, or perhaps even necessary as well (Newell and Simon 1976). It follows that its reproduction in computing machines will result in intelligence. On this classical view, artificial intelligence and cognitive science are just two sides of the same coin:

> Artificial intelligence is not the study of computers, but of intelligence in thought and action. Computers are its tools, because its theories are expressed as computer programs that enable machines to do things that would require intelligence if done by people. (Boden 1977: xi)

See also the classic Textbook: Artificial Intelligence: A Modern Approach, by Stuart Russell and Peter Norvig where they say at the outset "We define AI as the study of agents that receive percepts from the environment and perform actions." (Russell and Norvig 2010: viii). This expression has remained the same in the 1995, 2003 and 2010 editions. The only thing that was added in the latest edition is "We place more emphasis on partially observable and nondeterministic environments" (Russell and Norvig 2010: ix). Philosophically, the main thesis of classical computationalism is that the human mind is a functional computational mechanism operating over meaningful representations. These representations are caused by information-theoretical processes (Dretske 1981, 1995) or biological function in a "teleosemantics" (Macdonald and Papineau 2006; Millikan 2005). This account is motivated by classical 'machine functionalism, going back to (Putnam 1960) and nicely characterized by Churchland: "What unites them [the cognitive creatures] is that [...] they are all computing the same, or some part of the same abstract <<sensory input, prior state>, <motor output, subsequent state>>function." (Churchland 2005: 333). The set of functions that can be computed in this fashion is delineated by the 'Church-Turing thesis': All and only the effectively computable functions can be computed by a Turing machine–i.e. step by step, fol- lowing an algorithm (definite and finite rule). Machine functionalism together with a semantics make the basics for classical cognitive science and AI.

1.1 Critique of the computationalist picture

Of course, classical computationalism has come under criticism from many directions over the years, and some of that criticism has coincided with a perceived lack of tech- nical progress in technical AI.

We will not aim to give any details here, but allow me to mention a few mile- stones in that debate.

- Computation alone cannot generate intentional states of agents, especially the state of 'meaning something'. This problem has prominent forms in the 'Chinese room argument' (Preston and Bishop 2002; Searle 1980), the critique of 'encodingism' (Bickhard 1993; Bickhard and Terveen 1996), and others.
- The 'frame problem', one version of which seems to show that a computational system cannot make decisions without representing a very large number of facts (Dennett 1987; Pylyshyn 1987).
- Digital items like 'concepts', 'words' or 'phonemes' play little or no cognitive role, perhaps no representation plays much of a cognitive role (or none) - anti-representationalism and sub-symbolic cognition: (for example Bermúdez 2003; Calvo Garzón 2006).
- Human cognition presupposes a human condition (Dreyfus 1992–originally 1972; Wheeler 2005).
- Cognition is goal-dependent, thus a property only of certain biological creatures, that act–"enaction" or "deep embodiment" (Di Paolo et al. 2010; Froese and Di Paolo 2011).
- All and only the cognitive agents are embodied, cognition is largely a function of a body, etc. (Clark 1997, 2003; a useful introduction Hoffmann et al. 2010; Varela et al. 1991).
- ...

And from all this, one might conclude: "Cognition is not computation!"

[1] Anatolia College/ACT, Greece & University of Oxford, UK
Web: www.sophia.de

1.2 Two notions of computing:

At this point, we shall not discuss whether all these arguments and positions are any good. We will just try to clarify their impact, actual and argumentational. For these purposes, it might be useful to remember that there are two basic notions of 'computation' at stake here, that are fundamentally different:

- Computing over meaningful representations (GOFAI, 'cognitivism') (e.g. Dretske 1995; Fodor, 1998).
- Computing over meaningless syntactic tokens.

The failure to make this distinction has some pretty nasty effects:

Mistake 1: [Type: Throw baby out with bathwater]

> "Cognitivism is false, therefore cognition is not computation and AI [via computers] won't work."

Mistake 2: [Optimistic extrapolation]

> "I am not making Mistake 1, therefore cognition will still be computation and AI [via computers] with work."

2 Some Forms of Embodiment

Classical Embodiment

Allow me to expand on these forms of embodiment a little bit, to see the arguments Useful surveys are (Calvo Garzn and Gomila 2008) and (Shapiro 2011, 2012). Classical Embodiment is largely a negative thesis against Cognitivism and stresses the bodily experience:

> By using the term *embodied* we mean to highlight two points: first, that cognition depends upon the kinds of *experience that come from having a body* with various sensorimotor capacities, and second, that these individual *sensorimotor capacities are themselves embedded* in a more encompassing biological, psychological, and cultural context.
>
> ... sensory and motor processes, perception and action, are fundamentally inseparable in lived cognition. (Varela et al. 1991: 172f)

An evolutionary motivation (Wolpert)

"Why do we, and other animals, have brains? ... Now you may reason that we have one to perceive the world or to think. That's completely wrong! ... We have a brain for one reason, and one reason only, and that's to produce adaptable and complex movements. There is no other reason to have a brain." (Wolpert 2011)

Embodiment as offloading (Pfeifer)

Starting with an intuition against "Cartesian" centralized control, we try to design robots with simple local control, exploiting body dynamics and interaction with environment. This results in "The Emergence of Cognition from the Interaction of Brain, Body, and Environment" (Pfeifer and Bongard 2007; Pfeifer et al. 2007). The main illustrations are things like A) 'passive dynamic walkers' i.e. walking robots that need no energy, no motors, just walk down a slightly downward sloping surface exploiting their body dynamics in interaction with the properties of the environment (e.g. friction). B) Insect walking. For example a cockroach has ca. 1 million neurons, of which only 200 descending to body, so the walking movements of each of the six legs is not centrally controlled [Roy E. Ritzmann], but rather the result of locally controlled movement. C) Trout swimming–a trout can remain steady in a flowing stream by exploiting the eddies and whirls of the stream and of its own body with minimal or no energy use for sidewise movement (a dead trout can retain this position for some time). D) A host of robots that show complex behavior with little or no control, just due to their morphology. The 'Crazy Bird' robot with two constant motors but no sensors showed various behaviors with minor modifications of motor speed or leg friction[1].

An animal can thus walk over a rough surface by exploiting the elasticity of its body and reducing computation whereas "a robot built from stiff materials must apply complex control to adjust to uneven ground and will therefore be very slow." (Pfeifer and Bongard 2007: 97). This notion is (unfortunately) called 'morphological computation' but is really a non-computational aspect of intelligence (Mller and Hoffmann in preparation). One why this approach can only be part of the story is the inherent tension between the stability of morphology and the flexibility required for complex intelligent behavior.

Embodiment as enaction (O'Regan)

Perception in general and seeing in particular is a kind of action–and this explains 'how it feels' to us (O'Regan 2011). Since perception is a kind of action, it requires a body (not just passive sensors).

Cangelosi

There are a number of cases where embodiment influences cognitive processing in more or less surprising ways–thus discrediting the traditional 'Cartesian' view of cognition as totally detached from a body. One method in empirical research to bring out these influences is 'priming' and thus a detection of a bias. For example:

- Image recognition tasks: Subjects will press a button faster with the right hand than with the left if primed with images that suggest usage of the right hand–an object with 'affordance' to grasp with the right, e.g. a coffee cup with its handle on the right.
- If people are made to nod, they are much more likely to agree to a given statement.
- Priming with 'elderly' words (or slow animals) make people walk more slowly.
- Priming with rude words make people more likely to interrupt a conversation.

Embodiment as grounding (Steels)

Under the impression of arguments against intentional states (esp. meaning) in computational systems, Harnad formulated the 'grounding problem': "How can the meanings of the meaningless symbol tokens, manipulated solely on the basis of their (arbitrary) shapes, be grounded in anything but other meaningless symbols?" (Harnad 1990, 335). Researchers in robotics have worked on systems that provide this 'grounding' through interaction with the world, sometimes interacting with other robots and thus generating a 'convention' for symbols that denote objects or classes in their environment. Luc Steels has declared this problem 'solved' (Steels 2008), though tongue-in-cheek and I have my reservations (see below).

3 What was the argument, again?

It would not be totally surprising if at this point some confusion had set in, for what was supposed to be the argument, and for which conclusion? A number of candidates come to mind:

- Hardware matters for robot performance (more than we thought)

[1] http://www.eucognition.org/index.php?page=behavioral-diversity---crazy-bird

- Sensation is largely a kind of action (and thus needs a body)
- Large parts of the Brain deal with sensorimotor issues
- Representations need grounding
- Handicapped humans can't think (or think differently) [???]
- Computers without bodies won't be intelligent
  Clearly, this needs some clearing up and I want to suggest that there are three different kinds of theses in here, an empirical, a practical and a conceptual one.

3.1 Empirical,practical,conceptual

The theses are the following, in a first approximation:

- An *empirical thesis* (about natural cognitive agents, esp. humans). With the- se agents, it so happens that cognitive, affective and bodily aspects are inter- twined and inseparable (e.g. Ziemke @ EUCog 2011).
- A *practical engineering thesis* (on how to best make artificial agents with certain abilities); a thesis on 'cheap design'.
- A *conceptual thesis* (about the necessity of a body for cognition, or a particular body for particular forms of cognition).

4 Re-grouping: Non-cognitivist, non-embodied computing

To see how the opponents of these embodiment theses have re-grouped it is necessary to pick up on our point above that computing may be understood as a syntactic process (in fact, I think it must be understood in that way). A basic point is that computing as far as it is well understood is centrally digital computing, i.e. it operates on discrete states that are tokens of a type (e.g. the type '0'). The operations on these tokens are algorithmic, i.e. precisely defined step-by- step and 'effective' (leading to a result). This stands in a certain tension to classical computationalism–which Varela et al. call 'cognitivism', putting an emphasis on its notion of 'central processing' rather than on the form of this processing, namely computing.

4.1 Syntactical Computationalism

In order to motivate that there can be another form of computationalism, we need to explain a few things on the notion of computation.
Levels of description

A given computer can be described on three levels of description, and properties that it has on one level, it will typically not have on another. The levels are a) physical, in terms of its realization in a particular hardware with its physical properties; b) syntactic, in terms of its digital properties that can be described formally, and c) in terms of semantics, what the digital tokens represent, if anything.

Computational sufficiency

- At some functional level (perhaps several), the brain operates as a digital computer (syntactically, not over representations). This is sufficient to generate cognition.
- *Computational sufficiency thesis (CST)*: "... the right kind of computational structure suffices for the possession of a mind, and for the possession of a wide variety of mental properties." (Chalmers 1994, 2012; Shagrir 2012a, 2012b)

Computational universality

- *Church-Turing Principle* "Every finitely realizable physical system can be perfectly simulated by a universal model computing machine operating by finite means" (Deutsch 1985: 99)
- "... everything that the laws of physics require a physical object to do can, in principle, be emulated in arbitrarily fine detail by some program on a general-purpose computer, provided it is given enough time and memory." [the problem is how the brain generates new explanations] (Deutsch 2012)

Multiple realizability

Strictly the same computation can be realized on different hardware (on several lev- els) and the same hardware can realize different computations if interpreted different- ly (on several levels). Here is an example:

We have a logic gate with two inputs, one output. The output is of 5 V if both in- puts are 5 V, otherwise 0 V (based on Sprevak 2010). This computes AND: output of 5 V ('true') iff both inputs are 5 V ('true'). This also computes OR: output of 0 V ('true') iff at least one of the inputs is 0 V ('true'). Which function is computed de- pends on how this system is used, which interpretation of the voltages is preferred. So this is a many-to-many relation: strictly the same computation (e.g. OR) can be realized on different hardware, and the same hardware can realize more than one computation. This suggests the *multiple realizability thesis*:

If a system is not multiply realizable, then it is not computational.

4.2 Example I: Brain functionalism

One growing area where syntactic computationalism is used is the representation of brain function in purely computational terms. Here is a classical starting point from Christoph Koch's *Biophysics of Computation*: "The brain computes! This is accepted as a truism by the majority of neuroscientists engaged in discovering the principles employed in the design and operation of the nervous system." (Koch 1999: 1) And what does that mean? It is thought of a sequence of incoming data–encoding–computational operations–control of output and a very liberal notion of computing is at play here. Something "can be thought of as computation as long as it can be mapped on one or more mathematical operations that perform some useful function" ... if it is "actually being exploited by the organism" (Koch 1999: 2). His example is that a marble running down a hill computes the "local minimum of a two-dimensional energy function".

If this is the way to see things, then perhaps we could scan the brain and emulate in different hardware? Given *computational sufficiency* (due to *computational universality* or for further reasons) and *multiple realizability*, this should be possible!

We do know the 320 neurons of the notorious C. Elegans nematode but as Koch says "We have no idea what the 302 neurons are doing!" (Ch. Koch, talk 2011). Ef- forts are now under way by David Dalyrumple to generate a full simulation of this organism (`http://nemaload.davidad.org/`), which achieves very complex behavior with these neurons–including finding food, reproduction and some learning.

For humans, the task would be just a tiny bit more complicated, with ca. 64 billion neurons (plus glia cells, etc.), ca. 200 cell types, ca. 7000 connections each via a long dendric tree that can span across the entire brain (Deca 2012). But efforts to detect the 'human connectome' of these connections are now under way and the EU has just awarded one of the two huge FET Flagship projects (10 years, 1 billion) to a computational study of the whole human brain in the

'Human Brain Project'. Several authors expect that whole-brain emulation might be the fastest way to high-level AI because it seems to require essentially scientific 'grind' on a large scale, but not deep 'insight' into the complexities of human cognition (Kurzweil 2012; Sandberg 2013).

4.3 Example II: 'artificial general intelligence'

Some AI researchers see the time has come to return from technical and specialized AI to the original aim of a universal intelligence, not unlike the human one, an 'artificial general intelligence' (AGI). If one starts on the assumption that an intelligent agent is one that successfully pursues their goals in a given environment by selecting the right action, then a more intelligent one can do this in more environments–this kind of consideration provides a general measure of intelligence (Legg and Hutter 2007). In this vein, one can work towards AGI with machine-learning techniques that essentially optimize output, given certain sets of input (normally with probabilistic techniques). Despite the fact that the original model has some unrealistic assumptions (agent has infinite computing power, is not part of environment, is immortal), there are substantial projects underway that create such agents (like AIXI) (Hutter 2012; Visser et al. 2006).
Note: Problems of 'Action-Selection' Allow me to note that this apparently innocuous line of research makes one particular assumption that seems problematic from an embodied point of view, namely that intelligent agents solve a problem of 'action selection', of 'what should I do next?' This is the outcome of a "Model-Plan-Act" view of action (with "Intention-Belief-Desire" psychology), which is dubious, even for humans.

In fact, life and cognition are continuous; there is no 'next step'. What counts as "next action" depends on the granularity of description (e.g. raise foot vs. go to supermarket), so there is no "set of possible actions" (life is not like chess). In this ac- count, it must be decided what is relevant information and which beliefs must change with action–the 'frame problem' is coming back to us. As an illustration, note that many intelligent agents do not 'select actions' at all: This seems apparent in lower- level animals (a slug or even a cockroach), in certain non-classical designs for AI and in coupled embodied systems; e.g. a passive dynamic walker.
Syntactic approaches Of course, there are more important models along these lines, in particular dynamic systems theory (e.g. Johnson et al. 2008; Spencer et al. 2009) or the view of the brain as probabilistic prediction machine (Clark forthcoming 2012). The point here was just to indicate that this kind of position exists and that it is untouched by several of the classical 'embodiment' arguments–in fact the latter two are advanced as endorsing embodiment.

4.4 Grounding ('Weak embodiment')

There is one other way to re-group in the face of embodiment but maintain a classical research program: admit that the symbols in a computer are initially meaningless, but try to ground these symbols through interaction with the world. What the precise shape of the 'grounding problem' is and whether it has been solved is a long story (Müller 2011, forthcoming), but I suggest to make the following distinction between two grounding problems:
The easy problem of symbol grounding

"How can we explain and re-produce the behavioral ability and function of meaning [and other intentional phenomena] in artificial computational agents?" This is an empirical question and a practical question, where solutions to the one are definitely useful for the other. Often practical proposals in 'epigenetic robotics' have been said to shed light on the mechanism in humans (Cangelosi and Riga 2006). As we mentioned above, some argue that the problem has been solved and the suitably constructed computational mechanism acquires a semantic network in interaction with other such mechanisms(Steels 2008), but this is hardly universally accepted (Cangelosi 2009).

Proponents of "deep embodiment" would have to say that computational-robotic solutions are bound to fail. None of these systems have any intentional states, desires or goals because they don't have a life, in particular a precarious one. Thus, they do not have the right functional architecture, the right causal connections for symbol grounding (Di Paolo 2010).
The hard problem of symbol grounding

Even if this problem is solved, there might be a harder problem, namely "How does physics give rise to meaning and other intentional properties?" To solve this would require to reproduce not only behavioral ability and function but also the right inner mechanism in order to "Get the system to mean what it says". In humans, the experience of understanding is an elementary part of what we call 'understanding', which is why the Chinese Room Argument turns on the presence or absence of this experience. (This relies roughly on a Grice's analysis, which Searle shares, namely: To mean what I say is to try to get someone else to recognize my intentions of meaning something–which might be different from what my words mean (Grice 1957).) It should be obvious that the hard problem directly involves conscious experience, i.e. it involves solving Chalmers' 'hard problem' of consciousness (Chalmers 1996). This problem is untouched by evolutionary robotics.

Given this situation, my view is that we should return to the 'easy problem': "How can we explain and re-produce the behavioral ability and function of meaning [and other intentional phenomena] in artificial [mainly] computational agents?" This is not a philosophical problem but one that can be solved with cognitive science. If symbol grounding is necessary for cognition at some 'level', this problem must be solved in order to achieve artificial cognition at that 'level'.

5 Is Cognition like adding numbers or like growing apples?

5.1 Causal powers

Given multiple realizability, would reproducing the computation in a cognitive sys- tem reproduce the behavior? I don't think so. The reason is that features on the physi- cal and semantic levels of description are not necessarily reproduced–but these are crucial for the causal powers, i.e. the behavior.

Given that *hardware-dependent features are not computational* (and thus "morphological computation" is not computation) we cannot expect such features to be identical in different realizations. For example, if one realization produces a red light, another might produce a barrier down. To use a more general example: A computational model of an apple tree does not produce apples (but only apples* in the model).

Given that *semantics-dependent features are not computational* ("GOFAI com- putation" is not computation in my terminology), we cannot guarantee that these will be identical in different realizations either. If one realization produces a YES, another might produce a NO, depending on the interpretation of the output. (Note that this is not the same point as the one above concerning AND and OR, which concerned syntax.)

5.2 Purely syntactic structure may be just what is needed for cognition...

As we illustrated above, there is some hope to think that purely syntactic structures might be just what is needed for a successful account of cognition that allows for successful artificial cognitive systems. Perhaps the syntactic properties that are maintained across multiple realizations are sufficient? Perhaps any realization of 2 + 2 = 4 adds 2 + 2? Or is this the old fallacy where "=" means something to me, and I thus assume that it means something to the computing machine? In any case, the challenges for artificial cognitive systems will remain gigantic, even if embodiment is not as much a game-changer as some have thought (Gomila and Mller 2012).

Also, we need to remember Mistake 2: (Optimistic extrapolation) - "I am not making Mistake 1 [no computationalism, thus no computing], therefore cognition might still be computation and AI via computers will work". This is not a given, this needs to be established.

6 Conclusion

As far as the three theses are concerned, if we remember that they are logically indepenent–and this is usually forgotten–, then we can say:

- The empirical thesis is true
- The practical engineering thesis is true
- The conceptual thesis is likely false (i.e. syntax is often enough) ... unless it should turn out that symbol grounding (easy or hard) is necessary, and that is not implied by the truth of the empirical thesis above.

References

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

Bickhard, Mark H. (1993), Representational content in humans and machines, *Journal of Experimental and Theoretical Artificial Intelligence*, 5, 285-333.

Bickhard, Mark H. and Terveen, Loren (1996), *Foundational issues in artificial intelligence and cognitive science* (Amsterdam: Elsevier Science).

Boden, Margaret A. (1977), *Artificial intelligence and natural man* (New York: Basic Books).

Calvo Garzón, Paco (2006), 'Towards a general theory of antirepresentationalism', *British Journal for the Philosophy of Science*, 59 (3), 259-92.

Calvo Garzón, Paco and Gomila, Antoni (eds.) (2008), *Handbook of cognitive science: An embodied approach* (Perspectives on Cognitive Science, Amsterdam: Elsevier).

Cangelosi, Angelo (2009), 'The symbol grounding problems has been solved: Or maybe not?', *IEEE CIS Newsletter on Autonomous Mental Development*, 6 (2), 10-12.

Cangelosi, Angelo and Riga, Thomas (2006), 'An embodied model for sensorimotor grounding and grounding transfer: Experiments with epigenetic robots', *Cognitive Science*, 30 (4), 673-89.

Chalmers, David J. (1994), 'On implementing a computation', *Minds and Machines*, 4, 391- 402.

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

Chalmers, David J.(2012), 'A computational foundation for the study of cognition', *Journal of Consciousness Studies*, 12 (4), 323-57.

Churchland, Paul M. (2005), 'Functionalism at forty: A critical retrospective', *Journal of Philosophy*, (102/1), 33-50.

Clark, Andy (1997), *Being there: Putting brain, body, and world together again* (Cambridge, Mass.: MIT Press).

Clark, Andy (2003), *Natural born cyborgs: Minds, technologies, and the future of human intelligence* (Oxford: Oxford University Press) 229.

Clark, Andy (forthcoming 2012), 'Whatever next? Predictive brains, situated agents, and the future of cognitive science.', *Behavioral and Brain Sciences*, forthcoming.

Deca, Diana (2012), '2012: The Connectome, WBE and AGI', in Ben Goertzel; Joscha Bach and Matt Ikle (eds.),*AGI-12 (Lecture Notes in Artificial Intelligence (SLNAI))*; Berlin-Heidelberg: Springer), 41-49.

Dennett, Daniel C (1987), 'Cognitive wheels: The frame problem of AI', in Christopher Hookway (ed.), *Minds, machines, and evolution: Philosophical studies* (Cambridge: Cambridge University Press), 129-52.

Deutsch, David (1985), 'Quantum theory, the Church-Turing principle and the universal quantum computer', *Proceedings of the Royal Society of London, A* (400), 97-117.

Deutsch, David (2012), 'Creative blocks: The very laws of physics imply that artificial intelligence must be possible. What's holding us up?', *Aeon*, (03.10.2012).

Di Paolo, Ezequiel A. (2010), 'Robotics inspired in the organism', *Intellectica*, 53-54, 129-62.

Di Paolo, Ezequiel A.; Rohde, Marieke and De Jeogher, Hanne (2010), 'Horizons for the enactive mind: Values, social interaction, and play', in John Stewart; Olivier Gapenne and Ezequiel A. Di Paolo (eds.), *Enaction: Towards a New Paradigm for Cognitive Science* (Cambridge, Mass.: MIT Press), 33-87.

Dretske, Fred (1981), *Knowledge and the flow of information* (Cambridge, Mass.: MIT Press).

Dretske, Fred (1995), *Naturalizing the mind* (Cambridge, Mass.: MIT Press).

Dreyfus, Hupert L. (1992), *What computers still can't do: A critique of artificial reason* (2 edn.; Cambridge, Mass.: MIT Press).

Fodor, Jerry A. (1998), *Concepts: Where cognitive science went wrong* (Oxford: Oxford University Press).

Froese, Tom and Di Paolo, Ezequiel A. (2011), ' The enactive approach: Theoretical sketches from cell to society.', *Pragmatics and Cognition*, 19, 1-36.

Gomila, Antoni and Müller, Vincent C. (2012), 'Challenges for artificial cognitive systems', *Journal of Cognitive Science*, 13 (4), 453-69.

Grice, Paul (1957), 'Meaning', *The Philosophical Review*, 66, 377-88.

Harnad, Stevan (1990), 'The symbol grounding problem', *Physica* D, 42, 335-46.

Hoffmann, Matej; Assaf, Dorit and Pfeifer, Rolf (2010), Tutorial on Embodiment, ed. Vincent C. Müller (Tutorials on Cognitive Systems; `http://www.eucognition.org`: EUCog Network).

Hutter, Marcus (2012), 'One Decade of Universal Artificial Intelligence', in Ben Goertzel and Pei Wang (eds.), *Theoretical Foundations of Artificial General Intelligence* (Amsterdam: Atlantis Press), 67-88.

Johnson, Jeffrey S.; Spencer, John P. and Schner, Gregor (2008), 'Moving to higher ground: The dynamic field theory and the dynam-

ics of visual cognition', *New Ideas in Psychology*, 26 (2), 227-51.

Koch, Christof (1999), *Biophysics of computation: information processing in single neurons* (Computational neuroscience; New York: Oxford University Press) xxiii, 562 p.

Kurzweil, Ray (2012), *How to create a mind: The secret of human thought revealed* (New York: Viking).

Legg, Shane and Hutter, Marcus (2007), 'Universal intelligence: A Definition of machine intelligence', *Minds and Machines*, 17 (4), 391-444.

Macdonald, Graham and Papineau, David (eds.) (2006), *Teleosemantics: New philosophical essays* (Oxford: Oxford University Pres).

Millikan, Ruth Garrett (2005), *Language: A biological model* (Oxford: Oxford University Press).

Müller, Vincent C. (2011), 'The hard and easy grounding problems', *International Journal of Signs and Semiotic Systems*, 1 (1), 70.

Müller, Vincent C. (forthcoming), 'Which symbol grounding problem should we try to solve?', *Journal of Experimental and Theoretical Artificial Intelligence*.

Müller, Vincent C. and Hoffmann, Matej (in preparation), 'Conceptual remarks on morphological computation and implementation-independence'.

Newell, Arthur and Simon, Herbert (1976), 'Computer science as empirical enquiry: Symbols and search', *Communications of the Association of Computing Machinery*, 19 (3), 113-26.

O'Regan, Kevin J. (2011), *Why red doesn't sound like a bell: Understanding the feel of consciousness* (New York: Oxford University Press).

Pfeifer, Rolf and Bongard, Josh (2007), *How the body shapes the way we think: A new view of intelligence* (Cambridge, Mass.: MIT Press).

Pfeifer, Rolf; Lungarella, Max and Iida, Fumiya (2007), 'Self-organization, embodiment, and biologically inspired robotics', *Science*, 318, 1088-93.

Preston, John and Bishop, Mark (eds.) (2002), *Views into the Chinese room: New essays on Searle and artificial intelligence* (Oxford: Oxford University Press).

Putnam, Hilary (1960), 'Minds and machines', *Mind, Language and Reality. Philosophical Papers, Volume II* (Cambridge: Cambridge University Press 1975), 36285.

Pylyshyn, Zenon W. (1987), The robot's dilemma: The frame problem in artificial intelligence, ed. *Theoretical issue in cognitive science* (Norwood: Ablex).

Russell, Stuart and Norvig, Peter (2010), *Artificial intelligence: A modern approach* (3 edn.; Upper Saddle River: Prentice Hall).

Sandberg, Anders (2013), 'Feasibility of whole brain emulation', in Vincent C. Müller (ed.), *Theory and Philosophy of Artificial Intelligence (SAPERE; Berlin: Springer)*, 251- 64.

Searle, John R. (1980), 'Minds, brains and programs', *Behavioral and Brain Sciences*, 3, 417- 57.

Shagrir, Oron (2012a), 'Can a brain possess two minds?', *Journal of Cognitive Science*, 13 (2), 145-65.

Shagrir, Oron (2012b), 'Computation, implementation, cognition', *Minds and Machines*, 22 (2), 137-48.

Shapiro, Lawrence A. (2011), *Embodied Cognition* (London: Routledge).

Shapiro, Lawrence A. (2012), 'Embodied Cognition', in Eric Margolis; Richard Samuels and Stephen Stich (eds.), *The Oxford handbook of philosophy of cognitive science* (Oxford: Oxford University Press), 118-46.

Spencer, John P.; Perone, Sammy and Johnson, Jeff (2009), 'Dynamic field theory and embodied cognitive dynamics', in John P. Spencer; Michael S. C. Thomas and James L. McLelland (eds.), *Toward a unified theory of development: Connectionism and dynamic systems theory re-considered* (Oxford: Oxford University Press).

Sprevak, Mark D. (2010), 'Computation, individuation, and the received view on representation', *Studies in the History and Philosophy of Science*, 41, 260-70.

Steels, Luc (2008), 'The symbol grounding problem has been solved, so what's next?', in Manuel de Vega; Arthur Glenberg and Arthur Graesser (eds.), *Symbols and embodiment: Debates on meaning and cognition* (Oxford: Oxford University Press), 223-44.

Varela, Francisco J; Thompson, Evan and Rosch, Eleanor (1991), *The embodied mind: cognitive science and human experience* (Cambridge, Mass.: MIT Press).

Visser, Beth A.; Ashton, Michael C. and Vernon, Philip A. (2006), 'Beyond g: Putting multiple intelligences theory to the test', *Intelligence*, 34 (5), 487-502.

Wheeler, Michael (2005), *Reconstructing the cognitive world: The next step* (Cambridge, Mass.: MIT Press).

Wolpert, Daniel (2011), *The real reason for brains (TED Talks)*.