AISB Convention 2015
University of Kent, Canterbury

Proceedings of the AISB 2015 Symposium From Mental "Illness" to Disorder and Diversity: New Directions in the Philosophical and Scientific Understanding of Mental Disorder

Edited by Joel Parthemore and Blay Whitby

Introduction to the Convention

The AISB Convention 2015—the latest in a series of events that have been happening since 1964—was held at the University of Kent, Canterbury, UK in April 2015. Over 120 delegates attended and enjoyed three days of interesting talks and discussions covering a wide range of topics across artificial intelligence and the simulation of behaviour. This proceedings volume contains the papers from the symposium entitled *From Mental "Illness" to Disorder and Diversity: New Directions in the Philosophical and Scientific Understanding of Mental Disorder*, one of eight symposia held as part of the conference. Many thanks to the convention organisers, the AISB committee, convention delegates, and the many Kent staff and students whose hard work went into making this event a success.

—Colin Johnson, Convention Chair

Contents

Psychiatry and the poverty of subjectivity:

How phenomenology can contribute to the validation of categories of disorder

Anthony Vincent Fernandez[1]

Abstract. Psychiatry, and especially psychiatric classification, finds itself in a state of crisis. Recent criticisms have been leveled by patient advocacy groups, psychotherapists, and even psychiatrists (including the chairs of both the DSM-III and DSM-IV taskforces). Most notably, the National Institute of Mental Health (NIMH) announced—just weeks prior to the 2013 publication of the DSM-5—that it will primarily fund studies that *do not use* the DSM-5 categories of disorder. In light of the problems of classification plaguing the field of psychiatry, a number of phenomenologists (including Aho, Parnas, Ratcliffe, Sass, Stanghellini, and Zahavi) have argued that contemporary phenomenological research into psychopathology should be used to guide the project of reclassification. While I agree with this claim, I argue that these phenomenologists have failed to delineate among a number of domains of phenomenological research. And, in failing to make such distinctions, are unable to distinguish between those areas of research that can be used to validate categories of disorder, and those that cannot.

In order to remedy this issue in contemporary phenomenological psychopathology, I here propose three domains of phenomenological research—1) existential structures, 2) modes, and 3) traditions. The first is understood as the domain of phenomenology proper, and consists of the categorial characteristics of human existence (e.g. intersubjectivity, embodiment, situatedness, etc.). The second is understood as the study of the various modes of these categorial characteristics (the modes of Situatedness, for example, include

anxiety, boredom, joy, etc.). The third is understood as the domain of hermeneutics proper, but is often included in phenomenological studies. It consists of the framework of meaning that sediments throughout cultural and biographical developments, shaping what we see things *as* (e.g. people from different religious backgrounds will experience different objects *as* sacred, without actively interpreting the meaning of these objects).

1 INTRODUCTION

Since the 1980s, psychiatric classification has been dominated by the American Psychiatric Association's *Diagnostic and Statistical Manual of Mental Disorders* (DSM). However, the DSM-5, released in May of 2013, was the target of searing criticism from patient advocacy groups, psychotherapists, and even psychiatrists (including Robert Spitzer [1], chair of the DSM-III taskforce, and Allen Frances [2], chair of the DSM-IV taskforce). However, the criticism with the greatest visibility and most significant ramifications came from the National Institute of Mental Health (NIMH). Just weeks prior to the publication of the DSM-5, Tom Insel, head of the NIMH, declared in a public announcement that NIMH funding will be largely reserved for studies that *do not use* the DSM-5 categories of mental disorders [3]. Instead, most funding will be awarded for studies that support the new Research Domain Criteria (RDoC) project in its

[1]Dept. of Philosophy, University of South Florida. Email: `avf@mail.usf.edu`.

attempt to develop scientifically (i.e. neurologically and behaviorally) validated categories of disorder.[2]

The major concern held by Insel is that psychiatric research has failed to correlate the diagnostic categories of the DSM with neurobiological mechanisms. In other words, the symptomatically delineated categories of the DSM, drawing primarily on references to patients' lived experience (e.g. cognitive distortions, emotional disturbances, delusions, or hallucinations) and expressions of subjective experience in behavior (e.g. insomnia/hypersomnia, tearfulness, or hyperactivity), have not been adequately correlated with relevant changes in the brain. In order to remedy this issue, the RDoC project seeks to delineate preliminary research categories of disorder using only third-person observable data—including neurobiological data and certain kinds of behavioral data [4][5].[3] As currently formulated, studies of the lived world of subjects with psychiatric disorders will play no role in the delineation of the preliminary research categories that will be drawn up by the RDoC project.

While I share Insel's concerns over the disutility of the DSM categories, especially in regard to their failure to map onto neurobiological mechanisms, I believe that the RDoC and other projects aimed as reclassifying psychiatric disorders have been too quick to dispense with a phenomenological orientation. I argue not only that references to lived experience are conducive to the preliminary delineation of abnormal phenomena for neurobiological research, but also that phenomenological psychopathology (with its roots in the tradition of 20th century continental philosophy) is an invaluable tool for obtaining just such data.

What this amounts to is an argument over which kinds of research can contribute towards the project of creating valid categories of disorder. Philosophers and psychiatrists such as Robins and Guze [6], and Jablensky and Kendell [7] have outlined at least four kinds of validity, including construct, content, concurrent, and predictive.[4] In following with Jablensky and Kendell's breakdown of the various kinds of validity, a category of disorder has construct validity when it "is based on a coherent, explicit set of defining features"; it has content validity when it "has empirical referents, such as verifiable observations for establishing its presence"; it has concurrent validity when it "can be corroborated by independent procedures such as biological or psychological tests"; and it has predictive validity

when it "predicts future course of illness or treatment response" [7].

I argue that phenomenology can contribute directly to content validity by clearly describing the form of subjectivity and the lived world of a person with the disorder in question, and it can contribute directly to construct validity by differentiating one form of pathological subjectivity from another by clearly distinguishing essential from non-essential features of disorder. By offering rich descriptions of the disorders in question and by drawing clear boundaries around these disorders (at least in the cases where such boundaries exist), phenomenology can indirectly contribute towards the other forms of validity by supplying preliminary, symptomatically homogeneous categories that are more likely to correlate with specific psychological and neurobiological tests, as well as predict treatment response and course of illness.

My argument in this paper is presented in five parts. First, I review the work of the psychiatrist Gordon Parker and his colleagues in order to illustrate how close attention to subjective dimensions of disorder can lead to better systems of classification. Second, I review the recent literature on the role of phenomenology in psychiatric classification, focusing especially on the work of Josef Parnas and Dan Zahavi. In so doing, I bring to light some of the inadequacies in these accounts, showing that they fail to distinguish among a number of domains of phenomenological research, and thus among an array of different *kinds* of changes in subjectivity and human existence. Third, I draw on both historical and contemporary work in phenomenology and hermeneutics in order to delineate the three domains of phenomenological research. Fourth, I revisit each of these domains in light of the particular aims of phenomenological psychopathology, illustrating the kinds of pathological shifts that might be investigated in each domain. Fifth, and finally, I offer a preliminary sketch of how attention to these distinctions can lead to new psychiatric classifications with greater validity.

2 PSYCHIATRIC CLASSIFICATION AND THE POVERTY OF SUBJECTIVITY

In addition to the general criticisms leveled against psychiatry and the DSM, major depressive disorder (MDD) has found itself in the public spotlight following the publication of a number of popular books criticizing issues of classification, diagnosis, and treatment. Topics such as the pathologizing of normal kinds of sadness [8], the extremely low efficacy of anti-depressants [9], and the rapid rise in the number of people who meet the criteria for a diagnosis of MDD [10] have entered into public discourse, adding to the already marred reputation of contemporary psychiatry.

One researcher who has taken such criticisms to heart is the Australian psychiatrist, Gordon Parker. For over a decade, Parker has been pushing against what he calls the unitarian model of depression, which posits that depression is a single category of disorder that may differ along some dimensions (but in most cases is only considered to have one dimension—severity). His dissatisfaction with this model of depression led him to an article written by Kendell [11] that reviewed the historical ways of classifying depressive disorders. Drawing from these historical categories as well as his own research, Parker proposed three categories of depressive disorders (with the third category being

---

[2] It should be noted that the RDoC is not itself a system of classification. As Cuthbert and Kozack state, "It might better be termed 'an experiment toward classification.'"

[3] The place of behavior in this debate is a complex one, and I cannot say much about it here. Both the DSM and the new RDoC project rely heavily on observations of behavior. One important difference is that in the DSM-III and later editions, behaviors that show up exclusively—or at least primarily—in a single category of disorder are prioritized. In the RDoC, pathological or abnormal behaviors that show up across the boundaries of disorders drawn in the DSM are prioritized, primarily for the purpose of narrowing down avenues for further research on the neurobiological mechanisms behind such behaviors (rather than mechanisms behind certain categories of disorder, since it is these categories that the RDoC has put into question).

[4] It might be better to state that each of these aspects—rather than being independent kinds of validity—can be used to enhance or increase the validity of a category of disorder. However, this still leaves open the question of what validity itself *is*.

a catch-all for a diversity of depressive disorders that require further delineation) [12–14].

The classification he developed is hierarchical, with each subsequent level of the disorder incorporating the features of the level below it while including at least one additional feature. The three categories are, from the top to the bottom of the hierarchy, psychotic depression, melancholic depression, and non-melancholic depression. Non-melancholic depression is characterized simply by depressed mood (which is an admittedly ambiguous and likely heterogeneous symptom reference).[5] Melancholic depression is, in turn, characterized by "observable (and not merely reported) psychomotor disturbance" [14]. This characteristic, not being found in non-melancholic forms of depression, is an essential feature and a clear marker of melancholic depression. Psychotic depression, being the final category, includes depressed mood and psychomotor disturbance, as well as psychotic features, such as delusions or hallucinations.

Further research on the treatment efficacy and the neurobiological substrates related to these categories supplied evidence for their having a higher degree of validity than the DSM category of MDD [16]. For example, Parker and colleagues show that two-thirds of subjects who meet their criteria for melancholic and non-melancholic depression improve with anti-depressant drugs alone, while only one-quarter to one-third of subjects who meet their criteria for psychotic depression improve with the same treatment. Further data shows that anti-depressants have markedly higher efficacy for people with melancholic depression than for people with non-melancholic depression. Also, the addition of neuroleptics to anti-depressant treatment in the case of psychotic depression shows a marked increase in efficacy (beyond the rates for the same treatments when given to those with melancholic and non-melancholic depressions). Finally, psychotherapy proved beneficial only in non-melancholic forms of depression, having little primary effect on subjects with melancholic and psychotic depressions.

These findings, along with preliminary data pointing to distinct neurobiological substrates related to each category of depression, offer considerable evidence for the validity of Parker's hierarchical classification (at least when compared with the DSM category of MDD). However, what is most intriguing about this system of classification (at least for the purposes of this paper), is that its divisions and categorizations were originally made without reference to neurobiological causes, instead drawing primarily on subjective and experiential phenomena, such as depressed mood, delusions, and hallucinations.[6]

In spite of the success of Parker's categorization, it must still be kept in mind that these distinctions were drawn using a fairly superficial account of human subjectivity. This is not to say that the categories or divisions are illegitimate. Rather, I argue that such methods of categorization and classification can be markedly enhanced by traditions that have richer and more robust accounts of human subjectivity at their disposal.

3 CONTEMPORARY PHENOMENOLOGY AND THE RECLASSIFICATION OF DISORDERS

This basic line of argument has been offered by a number of contemporary phenomenologists and phenomenological psychopathologists [15,17–20]. While each of these authors has approached the possibility of using phenomenological research to inform psychiatric classification, I here focus primarily on a paper by Parnas and Zahavi entitled, "The Role of Phenomenology in Psychiatric Diagnosis and Classification," as it deals with the issue most directly.

In this work, Parnas and Zahavi aim to show how the tools and frameworks developed by the classical phenomenologists—including Husserl, Heidegger, and Merleau-Ponty—can help psychiatric researchers focus in on previously ignored (but often central) features of disordered subjectivity. They even go so far as to claim that "a search for a faithful *description of experience* must be considered as a necessary first step in any taxonomic effort, including attempts of reducing abnormal experience to its potential biological substrate" (2002, 137). They trace this idea back to Jaspers, who stressed the need for careful attention to experience, whether this is achieved by 1) observing "gestures, behavior, [and] expressive movements" in an attempt to perceive the meaning of such bodily engagements, 2) directly questioning or interviewing the subject, or 3) considering written first-person reports by the subject herself [21,22].

According to Parnas and Zahavi, phenomenology's major contribution towards the elucidation of psychiatric disorders stems from its account of the "essential structures" of subjectivity that were originally delineated by the classical phenomenologists. While there are numerous essential structures that might be discussed, they focus in particular on phenomenal consciousness and self-awareness; temporality; intentionality; embodiment; and intersubjectivity. These make up some of the core dimensions of phenomenological research, and the authors clearly illustrate how phenomenological research on each of these essential structures might contribute towards the project of re-classifying mental disorders.

However, because of the plethora of recent research in phenomenological psychopathology and the ensuing divergence of phenomenological frameworks and emphases among those working in the discipline, there is a different sort of clarification that is in sore need of attention. This is what I refer to as the *layers of phenomenological research*. The delineation of these layers does not amount to an alternative way of distinguishing among the essential features of subjectivity discussed by figures such as Parnas and Zahavi. Rather, all of these essential features are encompassed by just the first of three layers of phenomenological research.[7]

---

[5] See Stanghellini [15] for a phenomenological critique of the symptom of "depressed mood."

[6] Some kinds of psychomotor disturbance also fall into the category of experiential or subjective, but in this particular case Parker includes the qualification that it must be observable by someone besides the subject herself. As a result, this particular symptom does not technically count as experiential or subjective. Nonetheless, it does point, or refer, to an experiential phenomenon.

[7] The layers I sketch here were originally articulated in a paper with Giovanni Stanghellini. However, I here use slightly different terminology and draw the divisions in a slightly different manner. This is done in part because the original paper was written with a focus on Jaspers, while this paper focuses more directly on the philosophical tradition of 20th century transcendental phenomenology. However, it is also the case that I have realized that the characterizations of some of the layers in the earlier

4 LAYERS OF PHENOMENOLOGICAL RESEARCH

I refer to the three layers of phenomenological research as 1) existentials[8], 2) modes, and 3) tradition. These layers are related to each other in a particular manner, which is to say, they are not *merely* distinct. They are in some sense hierarchical, with the subject matter of each domain being a condition for the subject matter of the following domains (e.g. modes are modes *of* existentials). However, they can also be related in terms of degrees of particularity. The existentials that are discussed in the phenomenological canon are typically understood as applying to any and all human subjectivities, thereby being universal. Modes, on the other hand, tend to be available to all subjects, but at any time a subject finds herself only in particular modes. The term "tradition," on the other hand, may refer more directly to the subject matter of hermeneutics, but is also taken up in phenomenological studies of how culture and even personal narratives shape the way a world shows up to us. In this sense, existentials are the most universal, while tradition is the most particular.

4.1 Existentials

Existentials (sometimes referred to as "existential structures," "essential structures," or just "structures") comprise the first layer of phenomenological research, and are typically considered to be the subject matter of phenomenology proper. Phenomenology, with its Husserlian goal of discovering the *eidos*, or essence of the phenomenon under investigation, seeks out the necessary, universal, and invariant characteristics of human consciousness and existence. It is these characteristics that we call "existentials."[9]

Another important, but oft ignored characteristic of existential structures is that they are categorial. That is to say, existential structures are categories of characteristics of human existence. To take an example from Heidegger's *Being and Time*, the existential that he calls *Befindlichkeit*, typically translated as "situatedness," "affectedness" or even "sofindingness," refers to the fact that human beings always already find themselves situated in and attuned to the world. However, there are a variety of ways one can be situated in and attuned to the world. Situatedness, then, refers not to my particular way of being situated and attuned, but to the category that encompasses all the possible ways of being situated, such as through fear, anxiety,

wonder, or boredom. It is this categorial characteristic that is considered an existential.

4.2 Modes

Modes make up the second layer of phenomenological investigation, but they are not, strictly speaking, the subject matter of phenomenology proper. This is because modes are, by their very nature, contingent and variable. They do not make up essential, categorial characteristics of human existence. To continue the example above, I can be attuned and situated by fear, anxiety, wonder, or boredom. But the very fact that I can be attuned through a variety of moods means that no particular mood is part of my essential, existential structure.[10]

There are at least two ways modes can be approached in phenomenological research. First, they can be approached for their own sake, which is to say, a particular mode can be studied with the express purpose of learning more about that mode. An example of this kind of study is found in Heidegger's lecture on boredom, in which he conducts a lengthy phenomenological investigation of this mood for the express purpose of understanding the ways we can be bored, and the ways boredom shapes the meaning and significance of our world. Second, modes can be investigated for the sake of discovering characteristics shared by all modes included in a particular category. For example, in this same lecture course, Heidegger distinguishes among three different kinds of boredom based on whether they are directed towards an object, a situation, or disclose the world as a whole. While these distinctions were derived from a study of boredom, they proved useful in understanding moods in general, and in this sense his investigations were able to shed light on the existential structure of situatedness as a whole [23].

4.3 Tradition

Along with existentials and modes, phenomenological research often involves the study of what may be termed "tradition." This term is used throughout the phenomenological canon, receiving considerable treatment in Heidegger's early lecture courses, as well as in *Being and Time*. It plays an important role in genetic and generative phenomenology more generally, especially in Husserl's later works, such as *The Crisis of European Sciences and Transcendental Phenomenology* and "The Origin of Geometry" [24]. The term is typically understood in a broad sense, referring to one's "totality of presuppositions." There is a range of terms that are related to, or sometimes used as synonyms for, tradition. Some of these are facticity, thrownness, hermeneutical Situation, history, culture, and prejudice.

In addition to the myriad ways of referring to tradition, there are at least two reasons it is made the object of phenomenological research. The first, which we perhaps see most often in Heidegger's early works (but also in the works of

---

paper were not adequate to the task at hand, and needed to be updated and revised. The divisions and definitions of these layers are still, to some degree, a work in progress.

[8] Existentials are typically understood as the subject matter of phenomenology proper. In some cases they are referred to as structures, rather than existentials, but the term "structure" [*Struktur*] is used in a variety of ways, both within and amongst the works of each phenomenologist. In light of the possibilities for confusion that are opened up by the sometimes loose definitions of "structure," I have decided to use the narrower term, "existential."

[9] Heidegger often speaks of "ex-sistence" as a standing outside of, transcending or, simply, openness. Understood in this way, we can take "existentials" as categorial characteristics of human existence that play a role in the openness of the lived world, or the way in which the lived world is opened up and articulated for us.

[10] Besides moods, there are a number of other modes that have been discussed in the phenomenological literature. However, most of the classical phenomenologists fail to offer clear and careful definitions of existential structures and modes, so I rely on *Befindlichkeit* (situatedness) and *Stimmung* (mood) here because they offer the clearest distinction between existentials and modes in Heidegger's texts.

Husserl and Merleau-Ponty), is the explicit interrogation of our (mostly) tacit presuppositions that shape our interpretation of whatever the phenomenologist is interested in investigating. In *Being and Time*, for example, Heidegger engages in explicit interrogations of our presuppositions with respect to our concepts of time, truth, and being. In order to approach these concepts as phenomena—which is to say, as the proper subject matter of phenomenological research—we need to first make explicit the presuppositions that are at play in determining our everyday, scientific, or even philosophical conceptions of these phenomena. In the absence of such an interrogation—taken as a pre-phenomenological investigation, or an investigation conducted for the purpose of preparing for a phenomenological investigation proper—the phenomenologist risks (or perhaps risks more severely) falling back into illegitimate conceptions of the phenomena at hand, thereby failing to return "to the matters themselves."

The second way in which tradition is approached in phenomenological investigations is simply for its own sake, or for the sake of better understanding the form of the lived, meaningful world within which a person (or a people) finds herself. In this sense, one's totality of presuppositions is not made explicit for the sake of escaping the presuppositions and developing our concepts anew. Instead, these presuppositions are made explicit in order to better understand the meaningfulness of the world one resides within. While such investigations are neither phenomenologically preparatory, nor phenomenology proper, they have held a central place in the canon since the advent of genetic phenomenology.

An example of this latter kind of phenomenological study of tradition is found in the work of Iris Marion Young. In her essay, "Throwing Like a Girl," [25] she discusses the modes of feminine embodiment, but she also discusses the fact that such modes are tied up with a kind of tacit cultural background that shapes the meaningfulness of certain entities within our world or the kinds of meanings things have for us. As she explains, many women in the contemporary, western, affluent world have a sense of their bodies as fragile, weak, or even as an obstacle. The body is not actively interpreted in these ways, but simply shows up *as* fragile or weak in everyday experience. Young speaks of some of the biographical and historical conditions that led to such senses of the body, but this genealogical aspect is not particularly important here. Rather, what I wish to stress in Young's work is that the various modes of feminine body comportment that she outlines cannot be adequately understood without reference to the traditions in and through which one is able to come into contact with the world. In other words, in order to actually understand the form of one's lived world, we need to include an account of one's existentials, modes, and traditions.

5 LAYERS OF RESEARCH AND PHENOMENOLOGICAL PSYCHOPATHOLOGY

With the distinctions among these layers of phenomenological investigation clarified, we can reexamine them within the explicit context of phenomenological psychopathology. First, investigations into existentials, because their aim is to discover those characteristics of human existence that are considered necessary and universal, seem to have no place in

phenomenological psychopathology. Psychopathology is, by definition, concerned with those aspects of human existence that can and do change. If existential structures are, in fact, invariant, then rather than being the objects of study for phenomenological psychopathology, they might instead act as the background, or framework, within which phenomenological studies of disorders can be conducted.

However, taking such an orthodox stance ignores some of the major developments of 20th century phenomenology—specifically those of Merleau-Ponty [26]. Through his engagement with cases of subjects with severe psychiatric and neurological disorders, Merleau-Ponty came to doubt the absolute necessity of the existentials discovered and articulated by Husserl and Heidegger. By reassessing the case of Schneider, a WWI veteran who underwent profound changes in his perception and motility after being struck in his occipital lobe by a piece of shrapnel, Merleau-Ponty was able to show that phenomenology could not do justice to Schneider's disorder if it remained bound to the belief in absolutely necessary existentials, or structures of human existence. In order to adequately articulate Schneider's disorder, he had to appeal to changes in certain categorial characteristics of human existence that neither Husserl nor Heidegger would have allowed for.

Merleau-Ponty's insights fundamentally altered the kinds of investigations open to phenomenological psychopathologists. However, the distinction between this new layer of investigation and the layer in which we can examine modes is not immediately clear in light of Merleau-Ponty's work. In order to adequately express the difference between changes in an existential structure itself, and changes in the mode of an existential structure, I here briefly outline two ways in which phenomenologists might characterize certain forms of depression.

One account might characterize the affective dimension of depression as a severe change in ground-mood, which is a pre-intentional, world-disclosive affect or feeling. This account, because it refers primarily to certain kinds of moods, and the role that these particular moods play in the disorder, is a modal account of depression. That is to say, it portrays depression as a distinctive mode of finding oneself situated in and attuned to the world.

An alternative account might characterize the affective dimension of depression not as a particular mood, or mode of situatedness, but instead as an erosion of situatedness and attunement as a whole. In other words, depression can be characterized by a degraded or diminished capacity for being situated in and attuned to one's world at all. Such an account better explains the loss of meaning or significance in the world of the depressed person, as well as the lack of intense moods, degraded affect, emotional insensitivity to context, and even diminished capacity for sensory stimulation.

Both accounts seem to capture important features of the experience of being depressed. However, what is important to note here is that the former account characterizes depression as a particular mood, or mode of attunement, while the latter account posits a change in the category of moods as a whole or, in other words, a change in the existential of situatedness. This illustrates the difference between phenomenological studies of changes in existentials, and phenomenological studies of changes in the modes of these existentials.

With the distinction between existential and modal changes made, we can examine the role that tradition, or one's totality of

presuppositions, plays within the context of phenomenological studies of psychopathology. Cooper [27] offers one such example in the context of a criticism of phenomenology's role in psychiatric classification. In order to undermine the role of phenomenology in delineating categories of disorder, she considers the possibility of "masked depression," a condition that received considerable attention in the mid 20th century, but is still discussed to some extent today. These conditions are described as "depressions that do not make people feel depressed" [27]. As she explains, "Those who believe in masked depressions claim that cultural conditions can make it the case that certain individuals manifest depression in atypical ways. For example, in a society that sees sadness as unacceptable weakness, patients might instead report somatic complaints" [27].

Cooper argues that if psychiatric conditions such as masked depressions exist, then phenomenological investigations of disordered subjectivity are not particularly important for psychiatric classification (although she admits that there are a few cases in which such investigations might be useful). She is able to come to this conclusion because masked depression is meant to illustrate the possibility of disorders with a single cause manifesting—and being expressed—differently within different traditions or cultural contexts. In other words, the experiences of depression can differ in important respects (even to the extent that one might be said to *not* experience his own depression), and this is used to claim that phenomenology—understood broadly as any analysis of subjective experience—is of little to no use in such cases.

In contrast to such arguments, I believe the distinctions I have drawn among the layers of phenomenological research can be used to overcome such a criticism and show how phenomenology is sensitive, at least in principle, to the implications of cultural differences in the manifestation of psychiatric disorders. Insofar as phenomenologists are actually considered with making explicit and overcoming our traditional prejudices, or totality of presuppositions, they are not simply describing lived experience or offering an account of the way things seem or appear to us. In order to get at the changes in existentials and modes involved in a particular disorder, phenomenologists need to attend to the possible ways in which such a disorder might be misinterpreted. Such an investigations might involve detailed studies of cultural norms and prejudices, along with standard characterizations of disorders in the DSM and other psychiatric literature, as well as historical studies of the characterizations and classifications of disorders.

6 VALIDITY AND THE LAYERS OF PHENOMENOLOGICAL RESEARCH

In light of these distinctions, and the possibilities they open up for phenomenological research into psychopathology, we can return to the earlier discussions of phenomenology's role in the project of reclassifying psychiatric disorders with the intention of increasing validity. We can ask about which layers of phenomenological research contribute to the various kinds of validity, and especially towards the project of neurobiologically validating disorders. While it may be the case that research in all three layers can enhance validity, the primary contributions are likely to come from descriptions of the existential, and in some

cases modal, changes that comprise a particular kind of disordered subjectivity.

For example, modal investigations, such as phenomenological studies of the features of anxious moods or feelings in generalized versus social anxiety disorders, might enhance construct validity by showing that the moods and feelings associated with these disorders do not differ in any important respect. In this case, the only relevant distinction between the two kinds of disorders may be that the population diagnosed with social anxiety interprets large groups or social events as threatening or imposing. Because this account would characterize these two anxiety disorders as analogous in terms of modes, but dis-analogous in terms of traditions or tacit presuppositions, scientific research into the neurobiological correlates of moods and feelings may not need to distinguish between the two disorders in their investigations. However, psychotherapists may still find accounts of traditions and presuppositions relevant in order to change how people with social anxiety interpret and experience large groups or social events.

The history of phenomenology offers us even more evidence for the role that distinctions based on existential changes might play in the neurobiological validation of psychiatric disorders. As mentioned above, the possibility that such existentials, or existential structures, might be capable of changing (or even being absent) was not broached until Merleau-Ponty's *Phenomenology of Perception*. In this work, Merleau-Ponty takes Husserl and Heidegger to task for their transcendental assumptions that prove to be unjustified in light of the case studies of subjects with severe neurological disorders that he reexamined.[11] The fact that our only examples of such existential changes come from case studies of subjects with severe neurological disorders gives us reason to believe that other existential changes might also have relevant neurobiological correlates.

In sum, I have argued that phenomenology, specifically in the form of phenomenological psychopathology, is capable of offering accounts of disordered forms of subjectivity that can offer us preliminary categories of disorder that are likely to have greater validity than the categories currently available in the DSM. However, in order to properly engage in such a task, phenomenologists must be clear about the layers of their research.

Studies such as those discussed above can contribute *directly* towards enhancing both content and construct validity by supplying rich descriptions of disordered subjectivity, and by clearly distinguishing one kind of disorder from another by pointing out essential versus non-essential features of each disorder. Such clarifications can contribute *indirectly* towards other kinds of validity by offering symptomatically homogeneous categories of disorder that can then be used in neurobiological research, drug trials, outcome studies, and even psychotherapeutic interventions. While phenomenology may not be where psychiatry should end, it is certainly where it should begin.

---

[11] The particular example I have in mind is the case of Schneider, considered in detail in Part I of *Phenomenology of Perception*. However, Merleau-Ponty considers a number of other cases throughout this part of the text that may also prove useful as a model for phenomenological research into psychopathology.

REFERENCES

[1]   R. Spitzer, APA and DSM-V: Empty Promises | Psychiatric Times, (2009). http://www.psychiatrictimes.com/apa-and-dsm-v-empty-promises (accessed April 6, 2015).

[2]   A. Frances, A Warning Sign on the Road to DSM-V: Beware of Its Unintended Consequences | Psychiatric Times, (2009). http://www.psychiatrictimes.com/warning-sign-road-dsm-v-beware-its-unintended-consequences (accessed April 8, 2015).

[3]   T. Insel, Director's Blog: Transforming Diagnosis, (2013). http://www.nimh.nih.gov/about/director/2013/transforming-diagnosis.shtml.

[4]   B.N. Cuthbert, T.R. Insel, Toward the future of psychiatric diagnosis: the seven pillars of RDoC, BMC Med. 11 (2013) 126.

[5]   B.N. Cuthbert, M.J. Kozak, Constructing constructs for psychopathology: The NIMH research domain criteria., J. Abnorm. Psychol. 122 (2013) 928–937. doi:10.1037/a0034028.

[6]   E. Robins, S. Guze, Establishment of Diagnostic Validity in Psychiatric Illness: Its Application to Schizophrenia, Am. J. Psychiatry. 126 (1970) 107–111.

[7]   A. Jablensky, R.E. Kendell, Criteria for assessing a classification in psychiatry, Psychiatr. Diagn. Classif. (2002) 1.

[8]   A.V. Horwitz, J.C. Wakefield, The Loss of Sadness: How Psychiatry Transformed Normal Sorrow into Depressive Disorder, Reprint edition, Oxford University Press, Oxford ; New York, 2012.

[9]   I., Kirsch Ph.D, Irving Kirsch Ph.D.'sThe Emperor's New Drugs: Exploding the Antidepressant Myth, Basic Books, 2010.

[10]   R. Whitaker, Anatomy of an Epidemic: Magic Bullets, Psychiatric Drugs, and the Astonishing Rise of Mental Illness in America, 1 edition, Broadway Books, 2010.

[11]   R.E. Kendell, The classification of depressions: A review of contemporary confusion.pdf, Br. J. Psychiatry. 129 (1976) 15–28.

[12]   G. Parker, Beyond major depression, Psychol. Med. 35 (2005) 467–474. doi:10.1017/S0033291704004210.

[13]   G. Parker, Classifying clinical depression: an operational proposal: Discussion paper, Acta Psychiatr. Scand. 123 (2011) 314–316. doi:10.1111/j.1600-0447.2011.01681.x.

[14]   G. Parker, Classifying depression: should paradigms lost be regained?, (2000). http://ajp.psychiatryonline.org/doi/10.1176/appi.ajp.157.8.1195 (accessed April 8, 2015).

[15]   G. Stanghellini, Disembodied Spirits and Deanimated Bodies: The Psychopathology of Common Sense, 1 edition, Oxford University Press, Oxford ; New York, 2004.

[16]   G.S. Malhi, G.B. Parker, J. Greenwood, Structural and functional models of depression: from sub-types to substrates, Acta Psychiatr. Scand. 111 (2005) 94–105. doi:10.1111/j.1600-0447.2004.00475.x.

[17]   K.A. Aho, Depression and embodiment: phenomenological reflections on motility, affectivity, and transcendence, Med. Health Care Philos. 16 (2013) 751–759. doi:10.1007/s11019-013-9470-8.

[18]   J. PARNAS, L.A. SASS, Varieties of "Phenomenology," Philos. Issues Psychiatry Explan. Phenomenol. Nosology. (2008) 239.

[19]   M. Ratcliffe, A Bad Case of the Flu? The Comparative Phenomenology of Depression and Somatic Illness, J. Conscious. Stud. 20 (2013) 198–218.

[20]   L.A. Sass, J. Parnas, Explaining schizophrenia: the relevance of phenomenology, Reconceiving Schizophr. (2007) 63–95.

[21]   K. Jaspers, General Psychopathology, Reprint edition, Johns Hopkins University Press, Baltimore, 1997.

[22]   K. Jaspers, The phenomenological approach in psychopathology, Br. J. Psychiatry. 114 (1968) 1313–1223.

[23]   M. Heidegger, The Fundamental Concepts of Metaphysics: World, Finitude, Solitude, Indiana University Press, Bloomington, 2001.

[24]   E. Husserl, The Crisis of European Sciences and Transcendental Phenomenology: An Introduction to Phenomenological Philosophy, Northwestern University Press, Evanston, 1970.

[25]   I.M. Young, On Female Body Experience: "Throwing Like a Girl" and Other Essays, 1 edition, Oxford University Press, New York, 2005.

[26]   M. Merleau-Ponty, Phenomenology of Perception, 1 edition, Routledge, 2013.

[27]   R. Cooper, Psychiatric Classification and Subjective Experience, Emot. Rev. 4 (2012) 197–202. doi:10.1177/1754073911430139.

The juice is in the detail: an affordance-based view of talking therapies

Mark McKergow[1]

Abstract. The burgeoning interest in enactive paradigms of perception and cognition offers an opportunity to reconsider how we conceive psychotherapy – 'talking cures' as functioning. In the past many therapy modes have focused on the over-riding importance of giving insight to the patient; knowing what caused the 'illness' provides a solid way to deal with it. Over the past half-century, more pragmatic forms of therapy focusing on behaviour change through adjusted thinking (cognitive behavioural therapy) have become commonplace.

But what does it mean to 'change our thinking' from an enactive perspective? If perception and cognition are direct engagement with the environment, what is changed by a therapeutic conversation? One answer lies in the idea of affordances [1] – the relationships between features of the environment and the abilities of the animal/person to interact with them. Recent views of affordances as dynamic [2] make even clearer the ways in which these factors may change and evolve.

The paper compares an affordance based view with practical examples from solution-focused brief therapy (SFBT), where recent developments have pointed to the power of developing detailed descriptions of 'better futures' and 'past instances' [3]. In such detailed conversations, everyday and overlooked events such as hugging a loved one when they return from work can become significant possibilities for building recovery. The paper will show examples and how such detailed descriptions can develop new affordances for clients.

One key aspect is how these features emerge and are developed during the therapeutic conversation. Do they come from the therapist or the client? How can the therapist help the client develop new affordances that are relevant without intervening with their own ideas about 'what ought to happen'? The ways in which conversations about affordances can be seen to connect to strong and modest ideas of narrative development will also be explored briefly.

1 INTRODUCTION

In a symposium entitled 'Reconceiving Mental Illness', we are invited to think broadly about the topic. I intend to take this invitation seriously and present a novel view of both mental illness and how to enhance mental health. These topics have been discussed for centuries, and I cannot hope to present the full historical discussion here. Rather, I intend to set out some key points and then present a philosophical and practical case for a new way to look at mental illness though affordances.

One of the great truths (and for some, mysteries) of the mental health profession is that most if not all forms of talking therapy have broadly similar effectiveness. The huge metastudy of Wampold [4] showed that not only do different therapy modes have similar effectiveness, and drew attention to the overall importance of 'common factors' (first listed by Lambert [5]). These include therapeutic relationship/alliance, hope/expectancy, client factors and extraneous events. Despite this, the therapy world has continued to debate different models and approaches. One shortcoming of the Wampold study (and of most outcome studies) is the lack of consideration of the duration of therapy as of key interest. If everything 'works', then what works faster? During the heyday of psychoanalysis this was an unasked question, since it was common knowledge that mental disorders took years to deal with. During the past decades, however, there has been a rise in 'brief therapies', where the focus is on helping the client using 'as few sessions as possible' [6]. Such therapies typically take a handful of sessions to work [7].

There has been a bizarre obsession relating effective treatment to long-term therapy over the years, mainly due to the assumptions of psychoanalytic practitioners in the first half of the twentieth century. Clients and practitioners have grown more pragmatic in recent times, and now brief therapies are more valued. In a system such as the UK National Health Service where limited numbers of practitioners are available, the impact of shortening treatment can be huge. Lord Layard and colleagues [8] showed the huge impact of depression and other mental health problems – over a million people off work on incapacity benefit, in some cases waiting years to see a therapist who could help them in relatively short order (Layard mentions Cognitive Behavioural Therapy and 16 sessions). If the duration of therapy can be reduced from 16 sessions (itself brief by many standards) to closer to 4 sessions as shown by the latest brief therapy research [3], then four times as many people can be helped – even without recruiting extra therapists..

2 WHAT IS MENTAL ILLNESS?

This is a much contested question, about which there is little space to go into detail here. It looks so obvious at first, but unpicking the issues leads to considerable complication and confusion. The usual contrast is with physical illness – nobody would say that a broken leg was a mental condition. A stroke – a blood clot in the brain – can lead to speech impediments that can appear 'mental' (but probably should not be treated as such. Is pain mental or physical? Kendler [9] lists some of the key issues as causation (what causes mental illness, and in particular can it all be reduced to the brain, as some reductionists hope), the role of phenomenology and personal experience (which demands contact with the first person client situation rather than the third person expert) and nosology (the way that mental illnesses are classified). At present a pluralist view – different kinds of explanation are relevant – seems in the ascendant.

---

1 HESIAN, Department of Philosophy, University of Hertfordshire.

In general terms, most people think of mental illnesses as 'in the head'. One typical quote from the BABCP website [10] says:

"During times of mental distress, people think differently about themselves and what happens to them. Thoughts can become extreme and unhelpful. This can worsen how a person feels. They may then behave in a way that prolongs their distress."

This shows the assumption that thoughts precede behaviour – typical of the cognitive school of thought. This is so ingrained in our society as to go almost unchallenged – something 'inside' the person then appears on the 'outside' as behaviour. It is this assumption that the enactive paradigmn seeks to challenge.

3 THE ENACTIVE PARADIGM – DIRECT ENGAGEMENT

The enactive paradigm of perception and cognition is probably the most radical of the '4Es' [11] (embodied, extended, embedded and enactive) cluster of approaches which stem from the original work of Varela, Thompson and Rosch [12]. Briefly, rather than organisms taking in information ('perception') and then using it to make decisions about behaviour ('cognition'), the entire perception/cognition process is seen as a direct engagement with the environment. It is easy to see how this might happen for a blind person exploring a fruit bowl with their fingertips (or a pavement with their stick), but there are also indications and theories about visual perception based around sensorimotor rather than image-building processes [13].

Whereas the cognitive paradigm sees mechanisms in the head – either physical or mental – the enactive paradigm sees no need to posit mental representations. The world is its own representation, and carrying another around 'in our heads' would seem to be an unnecessary assumption. Indeed, Radical Enactive Cognition (REC)[14], the most extreme variety of enactivism, does away with all mental content. Another key distinction is the position of experience – our first person experience and awareness of what is happening to us. From a cognitive standpoint, experience is an epiphenomenon – a by-product of cognitive activity in the mind or brain (which are routinely superposed). In enactivism, experience is a primary element of cognition and is to be taken seriously in any description of 'mental' activity[15].

4 ROLE OF THE BRAIN – THE TASK/TOOL METAPHOR

This switch in emphasis can lead some readers to think that enactivism posits no role for thinking or the brain. This is of course incorrect. The brain is a vital organ, and removing it will seriously impede the thinking of the subject involved! Rom Harré's task/tool metaphor [16], [17] is a key way to understanding a way to look at the role of the brain from an embodied/enactive perspective.

Imagine somebody using a spade to dig a ditch. The person is using the spade to dig the ditch. The task is digging, and the tool , used by the person, is a spade. The spade does not dig the ditch – the person digs a ditch, using the spade. We could (and

should) study spades – after all, a well-designed spade will be a great help in digging the ditch. We can (and perhaps also should) study digging. Note the studying spades is not the same as studying digging, and to study digging we will need a person who is digging to make any progress in our study.

Now switch the task and tool to thinking and the brain. A person uses their brain to think. The person thinks, not the brain. We could (and should) study brains. However, to study thinking will require a person to do the thinking, in the same way that a study of digging requires a digger. To take on the idea that a brain thinks (as opposed to a person) is to commit what Maxwell Bennett and Peter Hacker call the 'meriological phallacy' [18] – applying to a part something which should only be applied to a whole. In this case the brain is a part of a person, and a person thinks (remembers, fears, loves, forgets, sees, etc), not a brain.

Memory can be treated the same way. Some people, including St Augustine [19] and Jerry Fodor [20] assume that memories must be treated like mental representations, carried around for reproducing at the desired moment. An enactive perspective makes clear that remembering is an activity of a person (not a brain), and involves an active constructive process – a re-membering, a putting together (as opposed to dis-membering, to pull apart). This view is being accepted in both scientific [21] and philosophical [22], [23] circles.

We might note that taking the task/tool metaphor seriously already offers a line on what constitutes a mental illness. One could imagine a separation between illnesses of the brain (for example brain tumours, strokes and even Alzheimer's disease) and diseases of the person (for example depression, anxiety). This is not to say that people are not incapacitated by brain diseases – far from it. It is interesting to note that Alzheimer's disease is formally classified as a mental illness in both the USA (within the DSM V [24]) and the UK (under the Mental Health Act 1983), which is probably a good thing in terms of sufferers getting practical help and protection under the law, but raises an interesting philosophical question.

5 IMPLEMENTATION

This paper promises an affordance based look at talking therapies. This section will take a look at affordances and the development of the idea over the past decades.

The term 'affordance' was originally introduced by ecological psychologist JJ Gibson [1], [25] in the late 1970s. Gibson's theory of direct perception, a precursor to the enactive paradigm, has three headlines:
- Perception is direct
- Perception is for action
- Perception is of affordances

Affordances are an interaction of an animal and its environment – what kind of opportunities for interaction the environment is offering the animal, relating to the animal's sensorimotor capacities. A small tree branch, for example, may offer a bird somewhere to perch and observe the surroundings, whereas the same branch might offer a person a handhold, a chance to gather kindling for a fire, a back scratcher, a drumstick, a subject for a sketch and so on. The affordance is neither a property of the animal or the environment, but in the interaction of both. Gibson himself defined affordances in this way:

[An] affordance is neither an objective property nor a subjective property; or it is both if you like. An affordance cuts across the dichotomy of subjective-objective and helps us to understand its inadequacy. It is equally a fact of the environment and a fact of behaviour. It is both physical and psychical, yet neither. An affordance points both ways, to the environment and to the observer. (Gibson, 1979, p 129)

Many people read Gibson as saying that the affordance is there to be discovered by the animal, in suitable ambient light. Varela, Thompson and Rosch [12] note that embodied perception is not 'direct detection' but is sensorimotor enactment, 'dependent on histories of coupling'. We might think of this as a learning process. Varela, Thompson and Rosch are also keen to emphasise the co-determination of animal and environment.

"A cognitive system is functioning adequately when it becomes part of an existing ongoing world (as the young of every species do." (p 207)

Anthony Chemero takes the idea of affordances on another level [2] with his 'affordances 2.0 model'. Having already refined his definition in an earlier publication [26] to be about the relationship between abilities of the animal and features of the enviroment (stressing further the learning element involved in developing affordances), he offers a dynamical model working on two timescales – developmental and behavioural. This shows even more clearly how abilities and affordances co-develop over both the life of an animal and over longer timescales.

Figure 1: Affordances 2.0 (after Chemero, 2009)

Sanneke de Haan, Erik Rietveld and co-workers[27] have further developed these ideas by contrasting the 'landscape' of affordances with the narrower 'field' of affordances for an individual in a concrete situation.

"We distinguish between the landscape of affordances and a field of affordances. The landscape of affordances refers to all the possibilities for action that are open to a specific form of life and depend on the abilities available to this form of life. In our human case this notably includes socio-cultural practices. The landscape of affordances thus describes the so-called "ecological niche" of a form of life. A particular aspect of the environment, say a tree, can play a role in the landscape of affordances of multiple forms of life. Von Uexküll (Von Uexküll, 1920)[28] gives the famous example of an oak tree: for a rabbit it affords digging a hole between its roots, to a woodworm it provides food, for a person it could afford shelter from sun or rain, or cutting. The field of affordances refers to the relevant possibilities for action that a particular individual is responsive to in a concrete situation, depending on the individual's abilities and concerns. The field of affordances is thus a situation-specific, individual "excerpt" of the general landscape of affordances." (from De Haan et al, 2013)

The phrase 'form of life' in this paragraph is a nod back to Wittgenstein's [29] adoption of this phrase to signify a context where language has a (shared) meaning. The authors then develop a three dimensional model to describe the extent of a field of affordances. The three dimensions are:

- Width (broadness of scope and choice of options)
- Depth (temporal – now and in the future, with anticipatory affordance-responsiveness)
- Height (relavance/important of affordances, relating to motivation and 'affective allure'

De Haan et al, who are seeking a way to describe the changes produced by deep brain stimulation treatment on sufferers from Obsessive Compulsive Disorder (OCD), tentatively sketch out how a field of affordances may appear in three different cases:
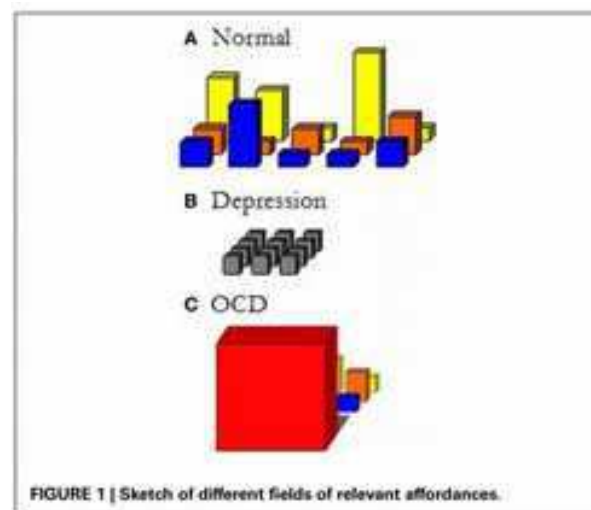
Figure 2: Sketch of different fields of relevant affordances (From De Haan et al 2013)

The 'normal' field of affordances show graspable variety in all three dimensions. The 'depression' verion shows everything looking the same – there is little in the way of meaningful choice or possibility of difference. The third diagram, reflecting the OCD case, shows one affordance (which may relate to washing hands or cleaning the house, for example) dominating the field in terms of importance. Note that these graphs are intended to be illustrative.

6 AN ENACTIVE VIEW OF MENTAL ILLNESS

German psychiatrist Thomas Fuchs offers an interesting way into a general discussion about enactivism and mental illness. In a paper [30] examining depression not as an inner and individual complaint, but as a detunement/disturbance ('Verstimmung') of 'the resonant body that mediates our participation in a shared affective' (which is very much stated in embodied and enactive terms), Fuchs harks back to phenomenologist psychiatrist Jan Hendrick van den Berg's pithy aphorism[31]: "*The patient is ill; this means, his world is ill.*"

Fuchs elaborates on this position: "In this sense, the illness is not in the patient, but the patient is in the illness, as it were; for mental illness is not a state in the head, but an altered way of being in the world". (Fuchs 2013, p 222)

Taking the statement 'the world of the patient is ill', it is easy and tempting to fall back into a cognitivist picture that the world of the patient is inside the head of the patient. From an enactive perspective, the world of the patient is 'out here', in the interactions of the patient. The recent developments in the theory of affordances described above now offer a way to expand on this idea in more concrete terms.

The 'world of the patient' is the patient's field of affordances. Remember that this is an excerpt from the total landscape of affordances open to the patient's form of life. This is dynamic on many levels – including behavioural and developmental. So, if we take those mental illnesses best described as conditions of a person (as opposed to a brain disease), we can tentatively define this form of mental illness as:

A persistent Verstimmung (disturbance/detuning) of a field of affordancesa

These terms are carefully chosen:

Persistent: Not very temporary – we all have temporary disturbances in our worlds and deal with them by everyday actions. We feel a bit miserable and decide to go out for a walk and see some friends, for example. These are everyday ups and downs, and are dealt with routinely most of the time. Only if the 'ordinary' ways of dealing with something prove ineffective can we start thinking in terms of illness. This idea was first put forward by John Weakland and colleagues at the Mental Research Institute, Palo Alto in the 1970s [32], [33] and is still sound.

Verstimmung: This is a German word which has a number of meanings difficult to entirely sum up in English. These include disturbance, detuning, and leaving a bad mood. This is not a breakage – there is a sense in which the disturbance can be corrected. This is not, of course, referring to a bad mood which 'accompanies' the illness, the Verstimmung is key to the whole picture.

Field: This refers to the field of affordances relevant to this person in this context. This inevitably brings a first person perspective into action – different people will naturally have different fields of affordance, and in particular the therapist/practitioner will not be able to take on the client's field of affordance.

Of affordances: This is, again, not in the person or the environment (though it is hard to speak of them in those terms with the limitations of English grammer, as in the paragraph above) but in the relationship between the person and their environment, as shown in possibilities for action and engagement.

7 AN AFFORDANCE BASED VIEW OF TALKING THERAPIES

Psychotherapy has been characterised (and caricatured) as 'two people talking, trying to figure out what one of the wants'. All talking therapies have in common at least the talking element (though the topics of the conversation very dramatically between approaches). We can also recall the findings of Wampold [4] that all talking therapies are about as effective as each other in pure outcome terms.

What has never been done, as far as I know, is to look at talking therapy explicitly in the way it stretches and changes the client's field of affordances. On this basis, therapies which seek to address mental distress by a focus on long-passed causalities such as childhood trauma and familial relations might be expected to take a long time to work, whereas therapies focusing more on details of the a better future might be expected to bring more rapid progress.

If we are to look at talking therapy as helping to stretch the client's field of affordances in useful ways that connect to progress, we might expect to look for:
- The therapist taking the client as an active participant in the treatment
- The therapist taking the first person perspective/descriptions very seriously
- The therapist not attempting to discover what has caused the problem, but rather establishing a conversational narrative around progress in the past, present and future
- The conversation being focused on small details of a 'better world' – signs that things were improving.

One might expect that such a stretching of the field of affordances might have an emergent quality about it – sometimes neat, sometime messy, sometimes clear, sometimes confusing. To stretch a field of affordances is not the same as to provide key steps for action to the client.

Might such a therapy be effective? Well, there is already one that works in much the above fashion which is indeed effective – Solution Focused Brief Therapy (SFBT).

8 SFBT THROUGH AN AFFORDANCE LENS

Solution-Focused Brief Therapy (SFBT) was devised by Steve de Shazer, Insoo Kim Berg and colleagues at the Brief Family Therapy Center in Milwaukee WI in the 1980s [34], [35]. It has since spread around the world, being widely used in education, social work, organisational change as well as therapy, with a significant evidence base [7], [36]. The approach appeals to those who value a pragmatic and skilful approach to building progress, but it has not been widely supported by psychiatrists and medical professionals for whom it lacks proper 'theoretical' grounding. De Shazer, Berg and colleagues started with the interactional brief therapy approach devised by Weakland and others, and experimented with trying to make it both more minimal (in terms of the therapist's model and theory) and more efficacious (in terms of fewer sessions to help clients reach a position where they could carry on under their own steam,

without continuing therapy). In this way, the practice could be said to be pragmatically and empirically rooted.

The latest and most stripped down version of SFBT is that proposed and practiced by the BRIEF group in London [3]. In a typical first session, the therapist will:

- Discuss 'best hopes' of the client for the work together – a theme for the project
- Elicit a description of a 'preferred future' – with these best hopes realised
  - o Tomorrow (usually)
  - o Detailed and observable (referent)
  - o From client's perception and relevant others' positions – spouse, colleagues etc
  - o Suppose… all about how it could be, not how to get there
- Elicit 'instances' – in the past and/or present – of the preferred future happening already
  - o Often using a scale from 1-10
  - o Details, details, details…

In follow-up session(s), the therapist will ask about 'what's better?' since last time, seek more details about how the client managed to do that, and summarise progress so far. Using this model, Shennan and Iveson report (over an admittedly small number of clients) an average therapy duration of under four sessions.

It is generally found in practice (by me and others) that getting these conversations down into small tiny details is important. SFBT co-founder Insoo Kim Berg used to advise therapists learning the approach to value '$5 words' (very small concrete and everyday words) over the '$5000 words' of abstraction and professionalism typically used valued by self-important experts. I want to put forward the idea that these details are connected with stretching the field of affordances.

9 A REAL LIFE EXAMPLE: MANDY AND THE CUDDLE

To give a brief flavour of an SFBT session, I include here a very short excerpt from a real conversation. 'Mary' (not her real name) has been referred for treatment following long term depression and suicide bids. This is her first session. The therapist (Chris Iveson of BRIEF) is in the middle of helping Mary to describe a better tomorrow, when an imagined miracle has realised her self-defined hopes of 'the past not pulling her back any more'. After about 25 minutes, they reach a point in the day when Mary's partner Jeff will return from work.

Therapist: And what is the first thing he would notice when he got home, even before you spoke? What is the very first thing?

Mary: I would be… instead of a worried, stressed, anxious look on my face maybe a smile.

Therapist: Okay. And what would be the first thing you would notice about his response even before he spoke?

Mary: I think my body language would just be so… you know normally he has to come looking for me whereas I would imagine that I would be open to go and cuddle him instead. You know? So…

Therapist: Would he faint or…?

Mary: Possibly, yeah, absolutely. You might have to have the paramedics on standby, yeah. I think it would be shock, but pleasant shock rather than shock shock.

Therapist: So where would that be? Where would you be cuddling him?

Mary: I would imagine that… because I do almost always hear him pull up. I never go to the door. I let him come in through the door and come find me. Whereas I would probably go find him.

Therapist: Okay, so that would be a different…

Mary: Yeah.

Therapist: And what would you notice about the way you cuddled him that fitted with this sense of peace and pleasure, of being you?

Mary: He describes sometimes that when he asks me for a cuddle… he said 'When I ask you for a cuddle…' and I do give it to him, he goes 'You are rigid and you almost… you cuddle me but you are pushing me away.' So I would imagine that it would be a much more natural, open embrace where I felt relaxed and safe enough to do that. Not rigid and tight.

Therapist: And what would you notice about his response to your cuddling and that kind of relaxed…?

Mary: I think that he would be delighted with how it felt to have a cuddle that didn't feel like he was a) having to ask for or b) being pushed away from.

Therapist: And what would you notice about his arms?

Mary: I think they might be quite tight around me and probably hold me for longer than normal.

Therapist: Okay. And what would you notice about how you handled that?

Mary: I think it would be quite difficult because you get so rehearsed in how you do things. Whether that be good or bad, that's how you are. So I think it would be quite a new experience to have that.

Therapist: And if you are feeling like hugging him?

Mary: Not wanting to let go either rather than wanting to break that embrace.

Therapist: Okay.

Mary: Because at the moment it's like 'Okay, cuddle, quick, out of the way.' Whereas to actually enjoy the embrace and feel it rather than just do it and break away from it.

Therapist: And what would you notice about him as you do eventually break away from the embrace?

Mary: I think that he would possibly be very happy to have experienced a… not always having to want to ask. To find… you know, for me to acknowledge his needs and be able to actually do that for him.

Therapist: And how would he know that you are pleased to have had that embrace? What would he notice about you?

Mary: Because I wouldn't be rushing away from him, looking at the next task that has to be done. It's like hugging Jeff is on the list, I've got to do that and then I've got to get on and do this and do that. I probably would maybe just stand there with him maybe and chat about his day rather than rush off and try and do something different.

Therapist: Is that when you might suggest a walk or would that be…?

Mary: After dinner maybe.

Therapist: After dinner? Okay. So what might you have for dinner?

Note that the therapist is not himself contributing to the details. He is rather asking questions which help Mary come up with her own details. He asks questions such as:

- And what is the first thing he would notice when he got home, even before you spoke?
- And what would you notice about the way you cuddled him…
- What would he notice about you?
- And how would you respond, when he did that?

These questions are all in the context of Mary describing a future (tomorrow) that is both utterly mundane and yet transformed by the realisation of her own hopes. She is stretching and changing her world in response to the therapist's questions – and because the talk is of a better future, the stretching is in a potentially useful direction. (We might note that many therapeutic approaches take a lot of time talking about what happens when the problem occurs or started, which might be stretching the world in an unhelpful way.)

For clarity, some of the affordances discussed in the excerpt above might be:

- The sound of Jeff pulling up as an opportunity to go and meet him.
- Jeff's appearance as an opportunity for cuddling in a particular way.
- The cuddle as a longer engagement rather than something to be broken off.

I say these 'might' be affordances in the conversation. We cannot say from a third person perspectives what are new or important affordances - we would have to ask Mary herself. And I am not saying that it's now simply a matter of Mary going and doing these things – her world has been stretched, her field of affordances altered, and now life will go on. It is only later that the impact will be clarified.

Previous versions of SFBT have focused on the conversation as a route to the therapist being able to establish tasks or actions for the client to help them 'do more of what works'. The latest thinking from BRIEF, the author [37] and others is that such a direct interventionist approach is unnecessary – either asking the client what they are minded to do next, or even simply leaving that out of the conversation altogether seems even more effective. It is worth noting that when the client's description is as detailed as the example above, all sorts of tiny actions and reactions have become possibilities in a revised world. This supports my hypothesis that the world-stretching is the key, rather than any post-rationalising that may go on between client and therapist (though such further conversation may strengthen the new world in some way).

10 TALKING ABOUT AFFORDANCES AND BUILDING AFFORDANCES

We might legitimately ask about the connection between describing affordances and creating/using them. From a cognitive standpoint there is all the difference in the world between talk and action. From an enactive standpoint, the difference is considerably reduced. In order to describe something, the client has to somehow put themselves into a different world. And once it's been described it can't be undescribed – echoes of the social constructionist idea of Ken Gergen that we carry around all our previous interactions as potentials for action [38]. There is even a view that from the first person perspective of the client, there is no fundamental difference between information through language and through visual and corporeal channels [39], [40]. There is no space to go further into this fascinating position here.

One point worth making in closing – how this position relates to a narrative perspective, itself a popular strand of therapeutic thinking and practice with similarities and differences to SFBT[41]. There are some who hold 'strong narrative' views that everything in life should be viewed in narrative terms [42]. Others, with whom I would align my position [43] take a more modest view, embracing the idea that narrative offers a useful view rather than an overarching mechanism. This is consistent with the task/tool metaphor for the mind, where discourse is a key but not exclusive element.

11 CONCLUSIONS

This paper has covered a great deal of ground very quickly and lays out a potential agenda for investigation. The key points are:
- Affordances offers a new perspective for talking therapies
- There is intitial evidence that this perspective is useful on a practical basis
- This may go some way to show why some therapies take a lot longer than others
- This perspective offers a researchable hypothesis for even more effective forms of talking therapy.

ACKNOWLEDGEMENTS

REFERENCES

[1] J. J. Gibson, The ecological approach to visual perception. Boston: Houghton Mifflin, 1979.
[2] A. Chemero, Radical Embodied Cognitive Science. Cambridge MA: MIT Press, 2009.
[3] G. Shennan and C. Iveson, "From solution to description: Practice and research in tandem," in Solution-Focused Brief Therapy: A Handbook ofEvidence-Based Practice, C. Franklin, T. S. Trepper, E. E. McCollum, and W. J. Gingerich, Eds. Oxford: Oxford University Press, 2011, pp. 281–298.
[4] B. E. Wampold, The Great Psychotherapy Debate: Models, Methods, and Findings, vol. 62. 2001.
[5] M. J. Lambert, "Implications of Outcome Research for Psychotherapy Integration," in Handbook of Psychotherapy Integration, J. C. Norcross and M. R. Goldstein, Eds. New York: Basic Books, 1992, pp. 94–129.
[6] S. de Shazer, Keys to solution in brief therapy. New York: WW Norton, 1985.
[7] C. Franklin, T. S. Trepper, E. E. McCollum, and W. J. Gingerich, Solution-focused brief therapy: A handbook of evidence-based practice. Oxford University Press, 2011.

[8] R. Layard, D. Clark, S. Bell, M. Knapp, B. Meacher, S. Priebe, L. Turnberg, G. Thornicroft, and B. Wright, "The depression report; A new deal for depression and anxiety disorders," London: LSE, 2006.

[9] K. S. Kendler, "Why does psychiatry need philosophy," in Philosophical Issues in Psychiatry: Explanation, Phenomenology and Nosology, Baltimore: Johns Hopkins University Press, 2008, pp. 1–16.

[10] "What is CBT?," British Association for Behavioural and Cognitive Psychotherapists. [Online]. Available: http://www.babcp.com/files/public/what-is-cbt-web.pdf. [Accessed: 04-Apr-2015].

[11]]M. Rowlands, The New Science of the Mind: From Extended Mind to Embodied Phenomenology. MIT Press, 2010.

[12] F. Varela, E. Thompson, and E. Rosch, "The embodied mind," Cogn. Sci. Hum. Exp., 1991.

[13] J. K. O'Regan and A. Noë, "A sensorimotor account of vision and visual consciousness," Behav. Brain Sci., vol. 24, no. 5, pp. 939–973, 2001.

[14] D. D. Hutto and E. Myin, Radicalizing Enactivism: Basic Minds without Content. Boston: MIT Press, 2013.

[15] M. McGann, H. De Jaegher, and E. Di Paolo, "Enaction and psychology.," Rev. Gen. Psychol., vol. 17, no. 2, pp. 203–209, Jun. 2013.

[16] R. Harré, "Tasks, Tools and the Boundaries of the Discursive," Cult. Psychol., vol. 7, no. 2, pp. 145–149, Jun. 2001.

[17] R. Harré and F. M. Moghaddam, Psychology for the Third Millennium Integrating Cultural and Neuroscience Perspectives. London: Sage Publications, 2012.

[18] M. R. Bennett and P. M. S. Hacker, Philosophical Foundations of Neuroscience. Oxford: Blackwell, 2003.

[19] Augustine, Confessions, 2nd edition. Indianapolis: Hackett Publishing, 2006.

[20] J. Fodor, Hume Variations. Oxford: Oxford University Press, 2005.

[21] D. L. Schacter and D. R. Addis, "The cognitive neuroscience of constructive memory: remembering the past and imagining the future.," Philos. Trans. R. Soc. Lond. B. Biol. Sci., vol. 362, no. 1481, pp. 773–86, May 2007.

[22] D. Moyal-Sharrock, "Wittgenstein and the Memory Debate," New Ideas Psychol., vol. 27, no. 2, pp. 213–227, 2009.

[23] E. Myin and K. Zahidi, "The Extent of Memory: From Extended to Extensive Mind," in Mind, Language & Action, Proceedings of the 36th International Wittgenstein Symposium., D. Moyal-Sharrock, V. Munz, and A. Coliva, Eds. Ontos Verlag, 2014.

[24] American Psychologyical Association, Diagnostic and statistical manual of mental disorders (5th ed.). Washington, DC: Author, 2013.

[25] J. J. Gibson, "The Theory of Affordances," in Perceiving, Acting, and Knowing, R. Shaw and J. Bransford, Eds. New Jersey: Lawrence Erlbaum, 1977.

[26] A. Chemero, "An Outline of a Theory of Affordances," Ecol. Psychol., vol. 15, no. 2, pp. 181–195, 2003.

[27] S. de Haan, E. Rietveld, M. Stokhof, and D. Denys, "The phenomenology of deep brain stimulation-induced changes in OCD: an enactive affordance-based model.," Front. Hum. Neurosci., vol. 7, p. 653, Jan. 2013.

[28] J. Von Uexküll, Theoretische Biologie. Berlin: Paetel, 1920.

[29] L. Wittgenstein, Philosophical Investigations. Oxford: Blackwell, 1953.

[30] T. Fuchs, "Depression, Intercorporeality, and Interaffectivity," J. Conscious. Stud., vol. 20, no. 7–8, pp. 219–238, 2013.

[31] J. H. van den Berg, A Different Existence: Principles of Phenomenological Psychopathology. Pittburgh PA: Duquesne University Press, 1972.

[32] P. Watzlawick, J. Weakland, and R. Fisch, Change: Problem Formation and Problem Resolution. New York: WW Norton, 1974.

[33] J. H. Weakland, R. Fisch, P. Watzlawick, and A. M. Bodin, "Brief Therapy: Focused Problem Resolution," Fam. Process, vol. 13, no. 2, pp. 141–168, Jun. 1974.

[34] S. de Shazer, I. K. Berg, E. Lipchik, E. Nunnally, A. Molnar, W. Gingerich, and M. Weiner-Davis, "Brief therapy: focused solution development.," Fam. Process, vol. 25, no. 2, pp. 207–21, Jun. 1986.

[35] S. de Shazer, Clues: Investigating solutions in brief therapy. New York: WW Norton, 1988.

[36] A. J. Macdonald, Solution-focused therapy: Theory, research & practice (second edition). Sage Publications, 2011.

[37] P. Z. Jackson and M. McKergow, The Solutions Focus: Making Coaching and Change SIMPLE. Nicholas Brealey, 2007.

[38] K. Gergen, "Relating with Self and Others," Interact. J. Solut. Focus Organ., vol. 5, no. 1, pp. 9–25, 2013.

[39] L. L. Barrett, Beyond the Brain: How Body and Environment Shape Animal and Human Minds. Princeton University Press, 2011.

[40] A. D. Wilson and S. Golonka, "Embodied Cognition is Not What you Think it is.," Front. Psychol., vol. 4, p. 58, Jan. 2013.

[41] M. White and D. Epston, Narrative Means to Therapeutic Ends. New York: WW Norton, 1990.

[42] A. Rudd, Self, Value, and Narrative. Oxford: Oxford University Press, 2012.

[43] D. D. Hutto, "Narrative self-shaping: a modest proposal," Phenomenol. Cogn. Sci., Feb. 2014.

Are mental disorders illnesses?
The boundary between psychiatry and general medicine

Valentina Petrolini[1]

Abstract. Are mental and physical disorders meaningfully comparable? Are we are entitled to characterize psychiatric disorders in terms of illnesses? Traditionally, most attempts to define what counts as an illness rely on some notion of *normal functioning* that has been altered or disturbed, where the "norm" is established from an evolutionary (Wakefield 1992; De Block 2008) or statistical perspective (Boorse 1977). In this sense, the substantial distinction between somatic and mental disorders may just reflect different stages of development within medical disciplines. In general medicine, clinicians have a clear idea of how organs normally function and thus can detect illnesses smoothly or with a small margin of error. The psychiatric case looks *prima facie* different: we currently lack an ideal model of brain functioning and the high variability among patients renders the diagnostic process particularly tricky. This argument reduces the distinction between psychiatry and general medicine to a *practical* matter. On this view, the high rates of misdiagnosis and disagreement among experts in the classification of mental disorders simply derive from a lack of knowledge about the brain (see Aboraya et al. 2006).

The main goal of this paper is to assess the argument above by showing that it stems from an overly simplistic conception of medical practice. On one hand, the diagnostic process in *general medicine* is not as straightforward as it initially appears, as some interesting studies on error and cognitive bias have recently shown. On the other, the core distinction between *psychiatry* and general medicine does not simply rest on practical issues: rather, the former exhibits some methodological peculiarities that are rejected by other disciplines within the medical field.

The paper is divided in four sections: in §1 I motivate the need for more theoretical precision in defining the notion of illness, making the case particularly compelling for psychiatry. In §2 I discuss some recent empirical studies on diagnostic error and cognitive biases in general medicine, and in §3 I evaluate whether these results can be meaningfully applied to psychiatry. Finally, in §4 I outline a *medical model* that aims at encompassing both somatic and mental disorders: in particular, I argue that in order to incorporate psychiatry within general medicine we need to adopt a *multi-level*, *holistic* and *dimensional* approach to illness.

1 THE NEED FOR THEORETICAL PRECISION

Within philosophy of psychiatry, the attempts to gain clarity from current definitions of *mental illness* have encountered a common difficulty. Psychiatry is a branch of medicine and thus a practical discipline whose main goals are to treat patients and alleviate suffering. As a result, not much work has been done to define concepts with theoretical precision, as suggested by the heated debate around classification and the DSM's new edition (see Cooper 2004 and Frances 2012). Consequently, among clinicians the question: "Is X a disease?" is often used as a shortcut for: "Should the person affected by X be subject to medical treatment?" This approach seems immediately problematic because doctors recognize that some conditions do not qualify as illnesses despite being treated (e.g. pregnancy or circumcision). Thus, the crude conditional: "*If* X needs to be treated, *then* X is a disease" should be discarded, at least because it does not reflect the common practice within medical sciences.

However, any attempt to define *mental illness* rests on having some conception of what counts as an illness in general: in this sense, the analogy between somatic and mental disorders becomes of paramount importance. On one hand, the two classes should be similar enough to be subsumed under the common label of "illness"; on the other, they should be different enough to motivate a principled distinction between the two sub-groups (see Brülde & Radovic 2006 and Brülde 2010). This network of similarities and differences between somatic and mental disorders has been extensively discussed both in the philosophical and psychiatric literature. For example, Culver & Gert (1982) attempt to draw the line by arguing that physical pain is "always *localized* to some part of the body" whereas mental suffering "is experienced by the *whole person*" (p. 89. Italics mine). Other authors – such as Boorse (1975) – adopt a more skeptical attitude by calling into question the validity of the analogy itself: "It seems an open question whether current applications of the health vocabulary to mental conditions have any justification at all" (p. 50). At the extreme of this spectrum, Szasz (1974) completely rejects the medicalization of mental disorders and argues that psychiatry should rather be concerned with "problems of living" – e.g. behaviors deviating from socio-cultural, moral or political norms.

1 Dept. of Philosophy, Univ. of Cincinnati, 2700 Campus Way, 206 McMicken Hall, Cincinnati (OH) 45221, USA. Email: petrolva@mail.uc.edu

Despite the difficulty to devise a precise definition, there are – at least – two reasons for advocating a more rigorous characterization of psychiatric illness:

a) The *social consequences* connected to a diagnosis of mental illness dramatically differ with respect to the ones connected to somatic ailments. Indeed, being classified as somatically ill presents a mixture of harmful and beneficial consequences for the patient (e.g. distress but also sympathy or support) whereas most mental disorders are still associated with various forms of stigma (e.g. shame, exclusion, discrimination). Since the personal and social implications of a psychiatric diagnosis may be highly disruptive for the patient, the highest level of precision would be needed in defining mental illness. This consideration becomes especially important in the light of Szasz's concerns about social control. For instance, equating "illness" with "in need of treatment" could allow psychiatrists to categorize all deviant beliefs and behaviors as mentally ill and thereby exercise some sort of coercive power over patients (see also Foucault 1964).

b) The identification of mental disorders also presents *legal* and *ethical* implications. For instance, most criminal systems do not rely on strict liability and thereby allow for excusing conditions (e.g. insanity). In the US, the *M'Naghten Rule* states that in order to successfully establish a defense on the grounds of insanity the party accused has to prove that – at the time of the crime – s/he was either not knowing the nature and quality of the act or s/he was not knowing that the act was wrong. Such a principle strongly connects legal and moral responsibility by acknowledging that no one should be punished for an action that was not committed voluntarily, but rather resulted from a "defect of reason" or a "disease of the mind" (see *M'Naghten Rule*). Again, these cases demand the highest level of precision: a sloppy characterization of mental illness runs the risk of unjustly punishing someone who should have been excused or applying the rule to someone who should have been convicted.

What a) and b) illustrate is that although instances of misdiagnosis in general medicine may have disruptive consequences (e.g. death of the patient), a lack of theoretical precision in psychiatry harbors implications that extend to the social, legal and ethical realm. Therefore, a more rigorous definition of illness that would comprise mental disorders is both desirable and called for.

2 DETECTING ILLNESS IN GENERAL MEDICINE: DIAGNOSTIC ERRORS AND COGNITIVE BIASES

From the discovery of bacteria to more recent microscopic and post-mortem techniques, diseases have come to be characterized in terms of "deranged biophysical structures, genes and molecules" (Kendell 1975, p. 306.). To this day, the most straightforward way to define somatic illness is by appealing to some form of *lesion* or structural *damage* of the body. This standard view raises three main issues: first, it relies on some notion of normal functioning that needs to be spelled out more or less precisely (e.g. prototypes). Second, it needs to account for individual variation while at the same time drawing a line to establish "where normality ends and abnormality begins" (Ibid., p. 308). Third, since not all deviations from the norm would be harmful – e.g. exceptionally high IQ – a distinction between positive, neutral and negative variations is needed.

Despite these potential problems with classification, general medicine seems to fare much better than psychiatry in terms of accuracy and reliability. The high rates of misdiagnosis and disagreement among psychiatrists support this point: for example, Kirk, Gomory & Cohen (2013) cite a recent estimate according to which the diagnostic error rate is 38% for ADHD and 21% for Oppositional Defiant Disorder (p. 170). The rationale behind this argument seems to be the following: *reliability* works as an indicator for the *validity* of a medical category, since a sound classification allows practitioners to distinguish between disorders and non-disorders in most circumstances. Due to the proliferation of false positives and false negatives, psychiatry's reliability appears tainted and consequently the whole classification of mental disorders is called into question. Yet, here I argue that the appeal to diagnostic unreliability *per se* fails to draw a meaningful distinction between psychiatry and other branches of medicine. To support this point, I discuss a growing body of literature focused on *error* and *accuracy* in various medical disciplines, showing that the diagnostic process – even for somatic disorders – is far from straightforward. These results are particularly interesting because they show that a complex array of factors – e.g. biases, modes of reasoning – can easily influence diagnosis. More specifically, *cognitive factors* are estimated to be responsible for the majority of errors: for example, in internal medicine 74% of the misdiagnoses appear to have such an origin (see Graber, Franklin & Gordon 2005).

In a recent study, Graber & Berner (2008) confirm that "diagnostic errors exist at non-trivial and sometimes alarming rates" (p. S6). The extent of incorrect diagnoses varies significantly according to the specialty, with perceptual disciplines – such as radiology – scoring lower (2-5%) and clinical ones higher (12-15%). Other important factors seem to be the context of stress or uncertainty that facilitates hasty decisions (e.g. emergency room), whereas the presence of a second opinion tends to increase accuracy. Yet, studies using standardized vignettes to enable comparisons across experts show that clinicians wildly *disagree* with one another, and sometimes "even with themselves when presented again with a case they have previously diagnosed" (p. S5). Another core issue seems to be the *lack of feedback*: most physicians regard diagnosis as a "'one-shot deal', […] a stand alone, discrete episode of judgement" rather than a process that stretches over time and can be refined or amended through multiple interactions with the patient (p. S34). In particular, doctors do not take advantage of autopsies as an opportunity to learn from past mistakes, although – on average – 25% of autopsies reveal new problems that were not suspected clinically (p. S5).

Graber & Berner also present a series of studies on the issue of *overconfidence*, arguing that it may significantly contribute to diagnostic error. The level of overconfidence can be measured through practical indicators such as the clinician's tendency to disregard decision-support resources even when they are available and easy to access (e.g. national clinical guidelines). Cognitive aspects – e.g. arrogance, excessive reliance on expertise – can instead be observed through the failure to elicit complete information from the patient and the biased interpretation of results. Interestingly, all the studies point to a systematic misalignment between the degree of confidence and the degree of correctness: "The level of physician confidence showed no correlation with their ability to predict the accuracy of their clinical diagnosis. […] The confidence level of the worst performers was actually higher than that of the top performers" (p. S8). Friedman et al. (2005) offer more results in support of the negative impact of overconfidence on diagnostic accuracy. This study measured the tendency to seek for external tools in the diagnostic process (e.g. computer-based support systems, advice from colleagues), finding again a correlation between high levels of confidence and errors. In a nutshell, overconfident physicians seem less likely to look for external sources to back up their decisions, thereby increasing the possibility of error.

Other studies focusing more specifically on *cognitive factors* (e.g. flawed reasoning, faulty data gathering, poor interpretation) have been carried out by Mamede and her collaborators (2008 & 2010). Great part of their work aims at drawing clearer distinctions between the modes of reasoning used by physicians when performing diagnoses. Apparently, doctors tend to switch back and forth between two alternative cognitive styles. On one side, *non-analytical reasoning* based on the recognition of similarities between "illness-prototypes" and the case under review; on the other, *reflective reasoning* based on the effortful and step-by-step analysis of specific features. Mamede et al. (2008) show that factors such as the perceived difficulty of a case can influence the way in which physicians approach diagnosis: for example, it is sufficient to tell them that other colleagues have failed to interpret the situation correctly to trigger the passage from non-analytic to reflective mode. In the experiment two groups of physicians were asked to work on the same case descriptions, but only one of them was primed to see the context as "problematic": as a result, this group spent more time on the diagnosis and displayed a significant increase in accuracy.

Another possible interpretation of this result – not discussed by Mamede – draws on the overconfidence studies just discussed: when cases are perceived as more difficult, the level of confidence may decrease and then lead to a more accurate assessment of the situation. In other words, knowing that a colleague has already failed in evaluating a case would attenuate overconfidence and force the physician to evaluate the context more carefully – e.g. spending more time on the diagnosis or taking alternative possibilities into consideration. This interpretation is consistent with the data presented by Mamede: in the contexts perceived as "non-problematic" the rate of confidence was higher and the level of diagnostic accuracy lower, whereas in the "problematic" cases the opposite occurred.

A more recent study (Mamede et al. 2010) uncovers the fact that experience with clinical cases similar to one another may trigger inaccuracy: indeed, physicians tend to perceive the diagnoses that come to mind more easily as correct even when they are not (*availability bias*). This bias also seems to get worse as expertise increases, suggesting again either a switch to non-analytical reasoning over time or the development of detrimental overconfidence. Like in the previous study, a combination of both factors might influence the diagnosis, since experience usually correlates with a greater number of cases encountered as well as with an increase in confidence.

These studies show that appealing to reliability to motivate a distinction between psychiatry and general medicine may be misguided. Indeed – contrary to most expectations – alarming rates of misdiagnosis and cognitive biases affect various medical disciplines in a similar way. Therefore, taking reliability *per se* as an indicator for validity does not create a meaningful contrast between psychiatry and other branches of medicine, since they all appear to have serious issues with diagnostic accuracy. Rather, it would be more fruitful to acknowledge that the lack of accuracy can be caused by different *kinds of factors*. Some of them may be mitigated or corrected without having to change the underlying structure of the discipline (e.g. biases, modes of reasoning); others may require a more profound revision of assumptions and methodology (e.g. faulty taxonomy). In this section I have shown that diagnostic issues in general medicine normally arise from factors of the first kind; in the next section I turn to psychiatry and argue that factors of the second kind are more pervasive.

3 DETECTING ILLNESS IN PSYCHIATRY: PRACTICING IN A MINEFIELD

The very idea of applying the results on cognitive biases and reasoning errors to psychiatry has generated a good deal of controversy. For example, Groopman's book on medical reasoning – *How Doctors Think* (2007) – purposefully excludes psychiatry from the discussion: "I quickly realized that trying to assess how psychiatrists think was beyond my ability" (p. 7). Moreover, despite the common complaint about the high rates of misdiagnosis in the field, the empirical literature on psychiatric errors is still quite small and the few exceptions tend to focus on other aspects of the practice (e.g. medication errors). Some researchers – such as Crumlish and Kelly (2009) – have attempted to counteract this tendency by arguing that the cognitive style employed by psychiatrists is not "esoteric" or "un-understandable" but rather similar to the one employed in other medical disciplines (p. 72). Others have defended a mixed approach, according to which psychiatric practice may commit errors that are common to other medical specialties but also faces a series of additional issues due to its unique patient population. For example, Cullen, Nath & Marcus (2010) point out that the peculiar features of psychiatric patients may have an impact on the "nature, prevalence and preventability" of the errors affecting them (p. 198). Interestingly, in this study *diagnostic* errors are the least commonly mentioned by

practitioners (9%), whereas *medication* errors account for approximately one-third of the total (34%) and *preventive* errors – e.g. failure to implement safety protocols – stand at the top (40%). Both medication and preventive errors are motivated by factors unique to the psychiatric setting, such as the lack of expertise in dealing with some extreme behavioral manifestations (e.g. violence, resistance to treatment) and various forms of stereotypes and stigma towards patients.

These data show that the topic of diagnostic reliability remains rather unexplored in psychiatry. Yet, the fact that diagnostic errors are both less reported and less investigated may indicate a more substantial difference between psychiatry and other medical disciplines. As Phillips (2014) put it: "You cannot detect error unless you have a reliable, valid method of making diagnoses. Since the diagnostic process is less certain in psychiatry than in general medicine, that will make the detection of error less confidant" (p. 75). One asymmetry arises from the fact that psychiatry does not avail itself of laboratory tests or biomarkers, and detects disorders almost entirely through clinical evaluations (e.g. structured interviews). Due to this unavailability of external resources to back up the diagnosis, psychiatry often lacks reliable methods to spot cases of under-reporting or over-reporting. For these reasons, the level of risk and uncertainty already connected to general medicine becomes higher in psychiatric practice, to the point that the diagnostic process "could be likened to a minefield" (Kapur 2000, p. 399). However, at this stage the problem might still be considered *practical*: for instance, the absence of laboratory tests and biomarkers may reflect the current *lack of knowledge* about brain functioning. Yet, reducing the difference between psychiatry and general medicine to a practical matter runs the risk of obscuring other important asymmetries. Most importantly, it assumes that psychiatry and general medicine already adopt a common *methodology* when approaching diagnoses.

According to Murphy (2006), this methodology can be summarized in a *medical model* exhibiting two characteristics: 1) The commitment to a view that sees disorders as *breakdowns* in normal processes of various kinds (e.g. biological, cognitive, affective, etc…). 2) The idea that any taxonomy of disorders should be constructed with the goal of uncovering underlying *causes*. In other words: "Diagnosis is causal. [It] is a matter of uncovering the causal antecedents of visible pathology" (p. 324). While this model accurately reflects what happens in most branches of medicine, in psychiatry neither 1) nor 2) are satisfied. With respect to 1), psychiatric classifications tend to characterize disorders in term of *distress* or *disability* but do not rely on normal human capacities that have been damaged or disrupted. Consequently, the recent editions of the DSM do not aim at uncovering malfunctioning mechanisms but rather at describing different forms of deviant behavior. As Kirk, Gomory & Cohen repeatedly stress, the symptoms that are supposed to guide clinicians in the diagnosis often re-state in different ways what the disorder is supposed to be about. The criteria for ADHD are a case in point: the attention-deficit part is spelled out in terms of "difficulty to sustain attention" or "easily distracted", while the hyperactivity part is characterized by

actions such as "often leaves seat" or "often on the go" (2013, p. 167). With respect to 2), the DSM rejects any investigation on the causal underpinnings of mental disorders and advocates a *descriptive* approach that attempts to be "neutral with respect to etiology" (DSM-IV-TR, p. xxvi). In short, the rejection of 1) and 2) brings about a classification of mental disorders that neither focuses on the normal processes that are being *disrupted* nor attempts to understand what *causes* the disruption itself.

Psychiatry's disavowal of the medical model seems problematic for at least two reasons. First, it renders impossible to bridge the current gap between psychiatry and general medicine because the two disciplines are endorsing radically different *methodologies*. On one hand, the DSM defends a symptom-based approach based on the description of syndromes and completely divorced from theories or hypothesis about underlying causes. On the other, general medicine operates by constructing models of normal functioning and by grouping illnesses together via causal factors. In this sense, the problem appears more *epistemological* than practical: although our current understanding of the brain's functioning may be limited, the classificatory system in place prevents us from garnering more knowledge about mental disorders. Second, the adoption of a merely descriptive taxonomy creates *paradoxical situations* that become apparent once we re-apply a similar system to general medicine. If diagnoses were based on symptoms only, we would end up grouping together all the patients sharing similar clinical manifestations: "We would classify together everyone who coughs as sufferers from 'cough disorder' and thereby miss the fact that someone who coughs may be doing so for a number of very different reasons" (Murphy 2006, p. 312).

4 FITTING PSYCHIATRY INTO THE MEDICAL MODEL

Murphy's discussion on classification aims at uncovering the fact that psychiatry still remains distant from a full-fledged medical model. Here I expand on his proposal by suggesting a theoretical framework that would facilitate the inclusion of psychiatry within general medicine. In particular, I argue that a characterization of illness able to encompass somatic and mental disorders should be *multi-level*, *holistic* and *dimensional*.

Multi-level. The main barrier that prevents psychiatry from adopting a causal taxonomy consists in the fact that we are still quite ignorant with respect to the etiology of mental disorders. Many authors have highlighted the difficulty to reduce mental disorders to *brain pathologies*: for example, Kendell (1975) describes psychiatric patients as "behaving in ways that alarm or affront other people" and "believing things that other people don't believe" (p. 305). Broome and Bortolotti (2009) stress a similar point: "It does not take an expert to recognize that someone is mentally ill, but how would one decide whether dopamine quantal size, functional MRI activations, or repeats of genetic polymorphism were abnormal in the absence of a disordered person?" (p. 38). These passages point to the fact that – in order to diagnose someone as mentally ill – we often make use of norms that go beyond the somatic sphere to encompass

socio-cultural and epistemic factors. In this sense, most psychiatric explanations would appeal to the disruption of norms on *different levels*: for example, a patient suffering from the Capgras syndrome may present both a neurobiological abnormality (e.g. dopamine dysregulation) and an epistemic one (e.g. abnormal resistance to contrary evidence). Moreover, it would not always be possible to establish the correct level of explanation in advance: whereas for some disorders a fully biological account might suffice (e.g. Huntington's disease), for others we may need to appeal to socio-cultural factors (e.g. anorexia).

A multi-level approach could also be extended to *general medicine*: indeed, somatic illnesses are often the result of a complex array of factors ranging from faulty genes to unhealthy lifestyle. Obvious examples in this sense would be type-2 diabetes or lung cancer, where biological causes interact with environmental ones. Thus, both psychiatry and general medicine could benefit from a multi-level approach to illness. From a diagnostic viewpoint, taking a diverse group of factors into consideration would enhance our understanding of the *causes* behind diseases. For example, the social pressure to resemble women on commercials might matter more than genetic predisposition in the explanation of some eating disorders. Similarly, living in a culture where smoking has a particular social value may put a certain group of people at high risk of developing lung cancer (see Goldade et al. 2012). From a therapeutic viewpoint, a multi-level account allows to abandon a strictly pharmacological approach and to tackle diseases from different perspectives: e.g. cognitive behavioral therapy (CBT) in psychiatry; diet and exercise in general medicine.

Holistic. If somatic and mental diseases are the result of multiple factors and can be understood only by appealing to different levels of explanation, it would be important to explore the dynamics between them. For example, some recent studies have suggested a correlation between schizophrenia and dopamine regulation (see Kapur 2003 and 2004), while others have investigated the high incidence of this disorder within specific sub-groups of the population – e.g. immigrants in conditions of social defeat (see Cantor Graee & Selten 2005). By adopting a multi-level approach we grant that both factors may be useful to explain the onset of schizophrenia: on the biological level, a disrupted process of dopamine release; on the environmental level, risk factors such as migration history or adverse social conditions. Yet, the interaction between the two levels remains unspecified: Does the environmental condition of social defeat directly influence dopamine regulation (*state interpretation*)? Or rather, are the individuals already affected by this brain abnormality more likely to develop schizophrenia (*trait interpretation*)? The endorsement of a *holistic* approach takes advantage of both interpretations without having to consider them mutually exclusive. On one hand there is good evidence that social and cultural habits can shape neurological structures in meaningful ways: for example, taxi drivers appears to exhibit enlarged posterior hippocampal regions with respect to controls who are not experienced in spatial navigation tasks (see Maguire et al. 2000). On the other, chemical imbalances in the brain can affect behavioral manifestations in a variety of ways:

the well-known correlation between serotonin levels and depressed mood is just an obvious example.

By adopting a holistic approach, we characterize illness as an emergent phenomenon in which biological and environmental factors are almost invariably influencing one another. More specifically, it may be possible to construct a spectrum indicating the degree of interaction between different kinds of factors in somatic and mental disorders. On one extreme we would find those diseases that emerge almost independently of environmental interaction (e.g. Down syndrome); on the other, those primarily caused by socio-cultural pressures (e.g. bulimia). An interesting consequence of this approach is that the distinction between somatic and mental disorders would somewhat collapse, because the unit of analysis would become the entire organism and its relationship with the environment. This proposal also allows considerable flexibility in classifying a condition as a disease: for example, sickle cell anaemia protects the organism from malaria and thus can be considered an adaptive trait in sub-Saharan Africa, and a serious illness in other environments. In other words, what is functional or dysfunctional cannot be established in a vacuum: "It is difficult to know whether a condition is pathological without considering the environment in which it occurs" (McGuire et al. 1992, p. 93).

Dimensional. According to Murphy, psychiatry can fit a medical model only by endorsing a categorical view of illness, where a condition results from multiple interacting causes but still qualifies as "a distinctive destructive process afflicting a system" (2006, p. 357). A couple of observations can be made in response to Murphy: first – although many illnesses are defined categorically – there are also conditions that arise as a consequence of meeting or exceeding a threshold (e.g. hypertension, diabetes or obesity). These processes are more or less "disruptive" but could hardly qualify as "distinctive": thus, sometimes general medicine treats illness as a condition diverging *quantitatively* – rather than qualitatively – from normal functioning. Second, there is good evidence that many psychiatric symptoms are widespread among the non-clinical population. For example, in a study conducted on 586 college students, 30 to 40% report to have experienced auditory hallucinations at least once in their lifetime, and almost half of these even once a month (see Johns & van Os 2001). Delusions are another interesting example, since they seem to lie on a continuum with other utterly irrational beliefs: thinking that your spouse has been replaced by an impostor does not seem distinctively different from believing that breaking a mirror would bring you seven years of bad luck.

Admittedly, regarding many mental disorders as dimensional would mean drawing the line between pathological and non-pathological with a certain degree of arbitrariness. Yet, it also allows a greater degree of flexibility and the opportunity to evaluate the context on a case-by-case basis. For example, we may want to be conservative in setting the threshold for psychopaths, due to the serious legal and ethical implications often connected to this condition. At the same time, we may decide to pay special attention to "high-risk" situations that need to be monitored or acted upon (e.g. students who regularly

experience auditory hallucinations). This last point seems consistent with what happens in dimensional somatic disorders: for example, if my blood tests report high cholesterol or high sugar level – even within the limits – the doctor may suggest a change in diet or life-style to avoid more problematic consequences. Therefore – despite Murphy's concerns – the endorsement of a dimensional approach sits comfortably with the medical model and promotes a more nuanced view of medical practice. Indeed, it shows that an important part of medicine consists in dealing with *chances* rather than *causes* and that the distinction between pathological and non-pathological may be a matter of degrees (see Gigerenzer 2008).

To sum up, I start by asking whether an analogy between somatic and mental disorders could be meaningfully defended. Then, I appeal to some recent studies on accuracy and cognitive biases to show that the core distinction between psychiatry and general medicine does not rest on the issue of *reliability*. Rather, the *symptom-based* approach currently endorsed in psychiatry is mostly responsible for distancing the discipline from the medical model, creating a gap between the ways in which mental disorders and other illnesses are diagnosed. Finally, I propose a *multi-level*, *holistic* and *dimensional* approach to illness that encompasses somatic and mental disorders.

REFERENCES

[1] Aboraya, A. et al. (2006). The reliability of psychiatric diagnosis revisited: The clinician's guide to improve the reliability of psychiatric diagnosis. *Psychiatry (Edgmont) 3*(1): 41-50.

[2] Boorse, C. (1977). Health as a Theoretical Concept. *Philosophy of Science* 44: 542-573.

[3] Boone, C. (1975). On the Distinction between Health and Illness. *Philosophy and Public Affairs* 5: 49-68.

[4] Bortolotti, L. (2009). *Delusions and Other Irrational Beliefs.* Oxford: Oxford University Press.

[5] Broome, M. R., & Bortolotti, L. (2009). Mental Illness as Mental: in Defence of Psychological Realism. *HumanaMente 11*: 25-44.

[6] Brülde, B. (2010). On Defining 'Mental Disorder': Purposes and Conditions of Adequacy. *Theoretical medicine and bioethics 31*(1): 19-33.

[7] Brülde, B. & Radovic, C. (2006). What is Mental about Mental Disorders? *Philosophy, Psychiatry and Psychology 13*: 99-116.

[8] Cantor-Graae, E., & Selten, J. P. (2005). Schizophrenia and migration: a meta-analysis and review. *American Journal of Psychiatry*, 162 (1): 12-24.

[9] Cooper, R. (2004). What is wrong with the DSM? *History of Psychiatry 15* (1): 5-25.

[10] Crumlish, N., & Kelly, B. D. (2009). How psychiatrists think. *Advances in psychiatric treatment 15*(1): 72-79.

[11] Cullen, S. W., Nath, S. B., & Marcus, S. C. (2010). Toward understanding errors in inpatient psychiatry: a qualitative inquiry. *Psychiatric quarterly 81* (3): 197-205.

[12] Culver, C. M. & Gert, B. (1982). *Philosophy in Medicine.* New York: Oxford University Press.

[13] De Block, A. (2008). Why Mental Disorders are just Mental Dysfunctions (and nothing more): Some Darwinian Arguments. *Studies in History and Philosophy of Biological and Biomedical Sciences 39*: 338-346.

[14] Foucault, M. (1964). *Madness and Civilization: A History of Insanity in the Age of Reason.* Random House LLC.

[15] Frances, A. (2012). "DSM-5 is a Guide, not a Bible: simply Ignore its 10 Worst Changes." *Huffington Post Science.*

[16] Friedman, C. P. et al. (2005). Do Physicians Know When Their Diagnoses are Correct? *Journal of General Internal Medicine 20* (4): 334-339.

[17] Gigerenzer, G. et al. (2008). Helping Doctors and Patients Make Sense of Health Statistics. *Psychological science in the public interest 8* (2): 53-96.

[18] Goldade, K. et al. (2012). Applying anthropology to eliminate tobacco-related health disparities. *Nicotine & Tobacco Research 14* (6): 631-638.

[19] Graber, M. L. & Berner, E. S. (2008). Overconfidence as a Cause of Diagnostic Error in Medicine. *The American Journal of Medicine 121* (5): S2-S23.

[20] Graber, M. L., Franklin, N., & Gordon, R. (2005). Diagnostic Error in Internal Medicine. *Archives of internal medicine 165* (13): 1493-1499.

[21] Grasso, B. C. et al. (2003). What do we know about medication errors in inpatient psychiatry? *Joint Commission Journal on Quality and Patient Safety 29* (8): 391-400.

[22] Groopman, J. E. (2007). *How doctors think.* Mariner Books.

[23] Johns, L. C., & van Os, J. (2001). The continuity of psychotic experiences in the general population. *Clinical psychology review*, 21 (8): 1125-1141.

[24] Kapur, N. (2000). Evaluating risks. *Advances in Psychiatric Treatment 6* (6): 399-406.

[25] Kapur, S. (2004). Psychosis as a State of Aberrant Salience: A Framework Linking Biology, Phenomenology, and Pharmacology in Schizophrenia. *American Journal of Psychiatry* 160: 13-23.

[26] Kapur, S. (2003). How Antipsychotics Become Anti-'Psychotic' – from Dopamine to Salience to Psychosis. *Trends in Pharmacological Sciences* 25: 402-406.

[27] Kendell, R. (2001). The Distinction between Mental and Physical Illness. *British Journal of Psychiatry 178*: 490-493.

[28] Kendell, R. (1975). The Concept of Disease and its Implications for Psychiatry. *British Journal of Psychiatry 127*: 305-315.

[29] Kirk, S. A., Gomory, T., & Cohen, D. (2013). *Mad science: Psychiatric Coercion, Diagnosis, and Drugs.* Transaction Publishers.

[30] Maguire, E. et al. (2000). Navigation-related structural change in the hippocampi of taxi drivers. *Proceedings of the National Academy of Science of the United States of America 97*: 4398-4403.

[31] Mamede, S. et al. (2010). Effect of Availability Bias and Reflective Reasoning on Diagnostic Accuracy among Internal Medicine Residents. *Journal of American Medical Association 304* (11): 1198-1203.

[32] Mamede, S. et al. (2008). Influence of perceived difficulty of cases on physicians' diagnostic reasoning. *Academic Medicine 83* (12): 1210-1216.

[33] McGuire, M. T. et al. (1992). Evolutionary biology: a basic science for psychiatry? *Acta Psychiatrica Scandinavica 86* (2): 89-96.

[34] Murphy, D. (2006). *Psychiatry in the scientific image*. Cambridge, MA: MIT Press.

[35] Phillips, J. (2014). Detecting Diagnostic Error in Psychiatry. *Diagnosis 1*: 75-78.

[36] Szasz, T. (1974). *The myth of mental illness: Foundations of a theory of personal conduct*. Harper-Perennial.

[37] Wakefield, J. C. (1992). The concept of mental disorder: on the boundary between biological facts and social values. *American Psychologist 47* (3): 373-388.

An encounter between Attachment Theory and 4e Cognition

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

Abstract.

A number of research questions arise from an encounter between the elements of 4e cognition and Attachment Theory. These include: (1) whether the Attachment Theory concept of Internal Working Models should be understood in terms of analogue representations more in line with embodied cognition, in addition to traditional cognitivist representations like linguistically mediated narrative measures of attachment meaning?; (2) are infant-carer dyads best thought of as environments of contextual embedding for infant cognition or as an arrangement where the carer can actually extend the infant mind?; and (3) are attachment phenomena best thought of in traditional representational terms or should the attachment control system be reframed in enactive terms where traditional cognitivist representations are: (3i) substituted for sensorimotor skill-focused mediating representations, (3ii) viewed as arising from autopoietic living organism and/or (3iii) mostly composed from the non-contentful mechanisms of basic minds?; A theme that cross-cuts these research questions is how representations for capturing meaning, and structures for adaptive control, are both required to explain the full range of behaviour of interest to Attachment Theory researchers. Implications are considered for future empirical and computational modelling research, and clinical interventions.

1 INTRODUCTION

The infant-caregiver relationship not only plays a central role in social and emotional development, but also in exploration and learning [3, 9, 10]. A traditional cognitivist approach to explaining these phenomena would emphasise internal information processing, located within the individual mind. So this approach in Attachment Theory would focus on what is or should be in the infant's head. A theoretical approach that keeps cognition within the infant is seductive because of its conceptual simplicity and because this approach is more easily implemented in cognitive models that focus on the creation and transformation of internal representations [18, 19, 20]. The elements of 4e cognition - viewing cognition as embodied, embedded, extended, and enacted - all reject or radically reconfigure traditional cognitivism [16]. Whilst the core ideas in Attachment Theory were set out by John Bowlby in a series of papers and books between 1958 and 1982 [2, 3, 5, 6], the elements of 4e cognition are more recently defined [16], but have many earlier conceptual antecedents [8, 11, 29].

How should Attachment Theory respond when viewed through the lense provided by 4e cognition approaches in cognitive science? And which elements of 4e cognition provide the best match for the requirements of a theoretical revision for Attachment Theory?

[1] Birmingham City University, UK, email: dean.petters@bcu.ac.uk
[2] SUNY, Stony Brook, USA.

Concepts from Systems Theory [8] as well as from Developmental Psychology, are key antecedents for contemporary Situated Cognition ([9] p 35). As Clarke notes:

> *"developmental psychologists were probably among the very first to notice the true intimacy of internal and external factors in determining cognitive success and change. In this respect, theorists such as Jean Piaget, James Gibson, Lev Vygotsky, and Jerome Bruner, although differing widely in their approaches, actively anticapted many of the more radical-sounding ideas now being pursued in situated robotics"* ([9] p 35)

The dialogic nature of the infant-mother relationship is exemplified by many types of interaction, including: the infant's active participation in co-operative games, the infant directing the mother's attention to acts by itself, use of objects as topics in infant-mother dialogues, and social and emotional referencing. The mutually contingent nature of these dialogues is demonstrated by experimental studies which perturb the contingency caregiver or infant responses, and in observational research of infant interactivity with depressed mothers [25]. Whilst Bowlby's formulation of Attachment Theory includes cognitivist constructs, like Internal Working Models (IWMs) and hierarchical plans, through which relationship patterns are represented internally, he was also inspired by Systems Theory [3], emphasising that an infant's main caregiver is the most salient part of the infant's environment. So Attachment Theory conceptualises infant-mother relationship as being between two active partners. Therefore, contemporary approaches from situated cognition can form a natural updating for Bowlby's systems approach, and may also help refocus cognitivist elements that Bowlby proposed within Attachment Theory.

The embodied approach views the body and physical world as the context or milieu" for cognition, rather than cognition conceived as the operation of disembodied algorithms [21]. So an encounter between Attachment Theory and embodied cognition asks how attachment representations should be conceptualised, and whether the cognitive component of Attachment Theory could then be *"augmented with the incorporation of bodily sensations, physiological responses, and analogue computations that rely on the physical substrate within the attachment control system"* [21]?

The hypotheses of embodied and extended cognition are competing theories in situated cognition that both give greater emphasis to the role that situations and context play in human cognition than traditional cognitivism. The extended approach is more radical, claiming that external supports become part of a person's cognitive apparatus. The embedded approach is still strongly anti-cognitivist, but sees cognition embedded in external support rather than constituted

of external structures. A key question is: whether attachment relationships can sometimes be conceived as extending cognition or are better thought of as embedding cognition?

The enactivist approach views psychological activity as occurring in the dynamic engagement between organisms and their physical and social context rather than within themselves [15]. The mind and subjective experience are not seen as inherent in, or arising from, the individual, but as emerging, from the interaction between organisms and their surroundings [15]. So another key question include: is enaction, rather than traditional forms of representation, a better way to think about how previous experiences mediates ongoing adaptive behaviour, and can the attachment control system be revised to act as an enactive "lived experiential structure" ([21, 28] p xvi)?

The intention in challenging Attachment Theory with recent ideas from 4e cognition is to revise rather than replace or reject it, and also see which diverse elements from 4e cognition can operate in 'joint purpose', motivating a progressive revision of a well established theory.

1.1 Bowlby formulated the attachment control system concept but did not specify it in detail

John Bowlby rejected psychoanalytic theory as a basis for explaining social and emotional development. Instead, he formulated a new explanatory framework by combining scientifically respectable ideas that originated across different disciplines. In his first presentation of Attachment Theory, in 1958, Bowlby provided an alternative motivational basis for attachment by replacing a psychoanalytic explanation based on Freudian instincts with a motivation framework based on ethological behaviours. Whilst this framework was too simple to explain different stages in attachment development it could be augmented further with other scientific concepts. In 1969, in the first volume of his Attachment Trilogy, Bowlby's theoretical revisionism included a much broader range of currently popular concepts, bound together in the attachment control system framework. So Bowlby's concept of an attachment control system replaced Freuds concept of psychical energy and its discharge ([3], p 18) and wove together constructs from: Piagetian theory; Cybernetics; Artificial Intelligence; and Systems Theory. He presented reflex behaviours and behavioral chaining of fixed action patterns as an example of a simple organizing principle for control systems, and hierarchical planning as much more complex and flexible ([3], p 76). Internal Working Models (IWMs) and natural language allowed higher level processes of integration and control. Then in the second and third parts of the attachment trilogy Bowlby invoked concepts from cognitive psychology. For example, he explained Freudian defensive processes in terms of selective attention ([6], chapter 4), and explained recall, reflection and potential internal conflict in self image in terms of the distinction between episodic and semantic memory ([6], p 61-64). Figure 1 shows illustrates how the 'theoretical borrowings' that Bowlby made changed with what were the prominent ideas of the day.

However, in none of his descriptions of the attachment control system did Bowlby set-down precise enough arrangements for how varied information processing elements might be organised in a running simulation. This is not a surprise, at the time that Bowlby formulated Attachment Theory, there existed no simulation technology to combine information processing elements such as ethological behaviours, IWMs and hierarchical plans within a single information processing architecture.
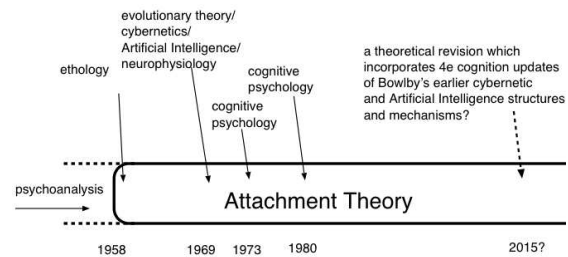
Figure 1. Diagram showing influences from other disciplines on Attachment Theory over time.

2 SHOULD INTERNAL WORKING MODELS BE VIEWABLE AS ANALOGUE IN ADDITION TO SYMBOLIC REPRESENTATIONS?

Internal Working Models are described by Bowlby as higher level representational forms which integrate and exert control over lower level control systems. Their principal information processing function is to allow predictions to be made about the likely outcomes of taking actions within a given environment. IWMs transmit, store and manipulate information and allow the individual to *"conduct small scale experiments within the head"* (([3], p 81). Their function, in terms of Bowlbys agenda of reforming psychoanalytic theory, was to take the place of the internal worlds of traditional psychoanalytic theory. Bowlby emphasizes the requirements for Internal Working Models to be updated. He also briefly observes that pathological sequelae of separation and bereavement can be understood in terms of out of date models or half revised models which may contain consistencies and confusions (Bowlby 1969 page 82). Bowlby invokes Internal Working Models at early stages in development also later on, when linguistic skills and conscious reflection can enable models to become more adequate ([3], p 84).

In contemporary reviews, IWMs are presented as transforming from sensorimotor representations in pre-linguistic infants to manipulable internal simulations in older children and adults that can enable short-term predictions, and conscious reflections on past, ongoing and future relationships ([7], p 102). Current research investigates IWMs through studies of memory talk, narrative completion, semi-projective measures and story-telling, with adults and children [7] - naturally linking IWMS to symbolic constructs from Artificial Intelligence like schemas and scripts. In his later writing Bowlby described IWMs in symbolic terms, for example:

"In reaching the decision to utilise certain actions rather than others the attachment system is conceived as drawing on the symbolic representations or working models, of the attachment figure, the general environment and the self, which are already stored and available to the system" ([4], p. 373).

However, links have also been drawn with IWMs and recent neuroscience research based upon mirror-neurons which presents IWMs as affording embodied simulation of the intentions of others ([7], p 109). Though research viewing IWMs as embodied simulations is very much in the minority in contemporary attachment research on IWMs ([7, 24], this section will argue that it is not only fully in the 'spirit' of Bowlby's original conception for IWMs, but also matches the 'word' of what he wrote about IWMs when he first introduced them. Bowlby did not use the term 'embodied simulation' but he did compare IWMs to analogue representations. For example, in his 1969 formulation of IWMs, Bowlby suggests that they can be used to

conduct 'small-scale experiments within the head' and notes that this notion would be an obvious possibility to electrical engineers familiar with analogue computers. Bowlby also refers to how anti-aircraft guns operate ([3], p 44) to exemplify how analogue control systems can set their own goals.

Looking back prior to 1969 to Bowlbys sources for the IWM concept provides added detail on how analogue representations can be conceived as mental models. Bowlby adopted the concept of Internal Working Models from the biologist J.Z. Young [31], whose treatment of Working Models is decidedly unambiguous in its preference for analogue over digital representations as a basis for Working Models in natural systems. As Young noted:

"[In an analogue computer] *the pattern of connections that determines what computation is made is part of the structure or pattern of the machine. These features at once suggest to the biologist, and especially the anatomist, that the nervous system is likely to work at least in part on analogue principles. What we commonly call the structure of the nervous system determines what it does. It is not a general purpose computer at all, but consists of a number of analogues set up to perform a few particular tasks. [] One of the great advantages of an analogue machine is that it can receive information directly from particular environments. That is to say, the machine maybe itself a representation of the environment and its parts are pre-selected to perform certain calculations in relations to the latter."* ([31], p 39)

J.Z. Young acquired the working model concept from its original source - the cybernetician Kenneth Craik. In The Nature of Explanation [12], Craik first discussed how working models can be used in science. Physical systems can act as models which help scientists explain natural phenomena because their physical operation captures key aspects of how the target system operates:

"*By a model we thus mean any physical or chemical system which has a similar relation-structure to that of the processes it imitates. By 'relation-structure' I do not mean some obscure non-physical entity which attends the model, but the fact that it is a physical working model which works in the same way as the process it parallels, in the aspects under consideration at any moment. Thus, the model need not resemble the real object pictorially; Kelvin's tide-predictor, which consists of a number of pulleys on levers, does not resemble a tide in appearance, but it works in the same way in certain essential respects*" ([12], p 51)

So in Craik's working models, although these systems can be argued to represent reality, when used by scientists to enable them to better explain and predict natural phenomena, it is by their physical properties rather than with abstract or arbitrary symbols that they represent other systems. Craik then made the significant leap to suggest that organisms can hold within their minds working models which operate in the same way. So living organisms can possess working models which represent their self and environment, and can run forward in time to make predictions or imagine the results of differing actions. Working models can also be configured to act as memories of past events.

The distinction between analogue and symbolic (discrete and digital) representations is important because analogue representations are much less flexible and are tied to the physical (embodied) properties of the medium in which they are implemented. Analogue systems

carry out computational operations using continuously varying data. Data in analogue devices is also transferred around these machines from input to output in continuous form and is bound to the physical form of the computational medium. So analogue computation relies on a physical or embodied substrate in a manner in which discrete symbol processing computations do not. These distinctions certainly matter to the growing number of researchers engaged in computational modelling of attachment behaviour, who actually want to implement running simulations of the attachent control sytems. In addition, how IWMs represent self and environment will also be of interest to clinicians who are concerned to activate, de-activate or transform attachment representations as part of therapy.

That Bowlby would invoke analogue computation and representations in his first formulation of IWMs might seem surprising given the contemporary predominance of the linguistic/symbolic approach to IWMs in Attachment Theory. It is in part explained by the waining popularity of analogue computers. In the period between the end of the second world war and the late 1960s when Bowlbys initially adoption of the working models concept, analogue computing remained a significant alternative to digital computing and the rise and domination of digital computing in the post-war years was not viewed as a foregone conclusion [27]. In addition, the seeming change in emphasis from analogue representations in 1969 to symbolic in 1982 may not represent a completely radical change in Bowlby's conceptualisation because Bowlby was vague in the representational details he proposed. As Bretherton and Mulholland note, Bowlby's formulation of the representational basis for attachment *"was a promising conceptual framework to be filled in by others"* ([7], p 103). However, perhaps the key issue was that in the 1960s Artificial Intelligence was less prominent in comparison with Cybernetics than it would be in the future. So the cybernetic view on issues like meaning and control held greater sway. This was consequential because researchers in Cybernetics under-emphasized representational distinctions and the challenges arising from consideration of high level processes. As Boden notes:

"*most cyberneticians seemed to see no difference between pure self-equilibration (as in homeostasis), purposive behaviour directed to some observable object (as in guided missiles), and goal seeking directed to some intentional end (as in human deliberation and planning)*"([1], p 220)

The eclipse of Cybernetics by Artificial Intelligence may have led to Bowlby's switch from invoking an analogue basis for IWM in 1969 to symbolic basis for IWMs in 1982. More recent developments have shown movement towards an integrative approach which might guide the process of bringing diverse representational forms together in the attachment control system, bringing back together a cybernetic approach to adaptive control and an Artificial Intelligence approach to fully intentional thought and reasoning [24, 22].

3 ARE INFANT-CARER DYADS BEST DESCRIBED IN TERMS OF COGNITIVE EMBEDDING OR COGNITIVE EXTENSION?

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, from the perspective of the socially situated 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 intimately linking internal infant cognition with external support [23]. So taking a situated cognition approach enrichs attachment theory by providing a more complete view of how infants gain information about environments from their caregivers.

Caregivers provide support to infant cognition in very many ways. They help to label, conceptualise, and structure information ([23, 10], p 44). Caregivers and infants are also situated within some of the same action loops that criss-cross close-coupled individuals and the environment [30, 9]. In such systems, caregivers can support 'soft assembly' of developing attachment competencies because secure attachment patterns are described in terms of response to set-goals rather than set actions ([9], p 44). Caregivers help scaffolding infant development by directing the child toward a correct/established outcome/solution/attitude or belief. When co-constructing they help the child take a course toward own-defined ends or end points. In addition, Bowlby describes how caregivers support infants by manipulating the environment and providing information directly through language use so 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"* ([3], p 82).

The hypotheses of embedded and extended cognition are competing and mutually exclusive explanations for how caregivers provide cognitive support. The hypothesis of extended cognition suggests that in some of the above examples, if the infant's ongoing computational needs are met by sensitive and timely support from his or her carer in such a way that the infant treats this support as part of their own cognitive processes then we might say that the carers cognitive support has become part of the infant's extended mind. For these examples to count as mind extension, caregiver cognitive support and information provision to the infant must be strongly trusted, relied upon and accessible. If these criteria are met then what is occuring is extension of mental states from an infant onto their caregiver. So in this view, the carer is actually extending the infant mind by incorporating the carer's help within the infant's cognitive operations - the carer's help becomes part of the infant's mind [3]. For these same examples of intimately integrated interactions between infant cognition and carer support, the hypothesis of embedded cognition views infant cognition and carer support of that cognition as clearly demarcated and separate. This hypothesis considers that *"cognitive processes depend very heavily, in hitherto unexpected ways, on organismically external props and on the structure of the external environment in which cognition takes place"* ([26] p 393). and that *"certain cognitive processes lean heavily on environmental structures and scaffoldings but not thereby include those structures and scaffoldings themselves"*([10], p 111).

We should be more accepting of claims to extended cognition in infants and younger children, because the caregiver's interactions are more long-lasting, they are relied upon more, and when there are less infant cognitive resources and routines for not believing [13]. So making acceptance of information from the carer as if it were an infant's own beliefs easier and more likely.

Two main reasons for preferring embedded explanations over extended explanations arise from considering non-social cognitive extension [10]. Most examples of extended cognition involve inorganic objects in the environment (such as a mathematician doing their 'working' on paper) providing the cognitive extension. The first criticism of extended cognition highlights the profound differences that appear to distinguish inner and outer contributions in extended cog-

---

[3] [23] presents a more detailed case that the infant carer dyad is an examplar of extended mind cognition, with the infant's cognition extended by their caregiver.

nition when cognition is extended onto such inorganic objects [10]. However, this criticism is much weaker when applied to the social case as it is a carer that does the extending. So there are not such profound differences in the supporting substrate for cognition between cognition inside the infant's brain and cognitive support originating from inside the carer's brain. A second criticism is the apparent scientific cost of any wholesale endorsement of extended cognition onto a motley collection of inorganic objects because it gives undue attention to transient external props and aids. In this view, following the extended mind hypothesis means scientists are not researching a suite of integrated persisting organismically grounded capacities [10, 30], and looking at developmental examples of cognitive extension onto inorganic objects is a series of separated developmental segments with external cognition onto different objects. So using a ball or balance beam may be a good example of mind extension at one age, but a year later the best example may involve a completely different object in a different task or action. Again, the social case of mind extension mitigates this criticism. Extended cognition does not only deal with transient external props and aids when the carer provides enduring support and continuity between otherwise disparate contexts.

If we accept the hypothesis of extended cognition over the hypothesis of embedded cognition this has important implications for computational modelling and in clinical interventions. Caregiving relationships are often very durable and reliable and if socially extended cognition occurs we can expect typical interactions and development to include micro and macro instances. Micro extension effects are described by Clark: *"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"* ([10], p67). Macro effects occurs when children absorb complex ideas wholesale through the conduit of cognitive extension. Their caregivers can simply present beliefs which the children then adopt. Over the long-term caregivers attempt to socialise and indoctrinate infants in many ways that will impact the developing meaning a child gains of their attachment history. Two types of problems can occur: (1) relationships are not reliable or durable enough so infants and children do not gain the benefits of cognitive extension; and (2) pathological extension occurs, so instead of acting to scaffold or co-construct, a caregiver uses their power to extend an infant's mind to introduce (or put more strongly 'infiltrate' or 'hack' [17]) unhealthy or pathological beliefs about the infant's self and relationships into the infant's mind.

4 ENACTIVISING ATTACHMENT THEORY

Where the extended/embedded question highlighted the requirement for attachment structures and mechanisms that support narrative meaning making the three flavours of enactivism highlight different aspects of adaptive control and subjective experience in the attachment domain.

4.1 Attachment Theory encounters Sensorimotor Enactivism

Sensorimotor enactivism criticises the view that perception results in inner images or mental representations being produced. In the sensorimotor view, perception, action, and subjective perceptual experiences are all inescapably connected [14]. This approach allows that perceptual experience is grounded in knowledge and is therefore

representationally contentful. But the kind of mediating knowledge in sensorimotor enactivist accounts is more like procedural or skill-based knowledge. It is 'know-how' rather than 'know-that', a kind of knowledge demonstrated by the skilled performance of its deployment rather than an independently queriable knowledge base [14].

Viewing attachent behavioural patterns in this enactivist manner - as social skills rather than arising as a result of internal representations - may provide a powerful spur towards new research hypotheses and clinical interventions. When individuals with insecure attachment gain secure status they can be viewed as gaining a skill which they can then use in other relationships.

4.2 Autopoeisis and representation from social interaction

According to autopoietic enactivism, cognition, mentality and subjective experience all emerge from the self-organising and self-creating activities of autonomous entities [14]. This activity is intimately spread between organism and environment. Enactivisits suggest that, because factors from 'within' and 'without' play equally important and necessary roles in creating cognition and behaviour, the distinction between organism and environment is viewed as only having a heuristic value rather than being a true metaphysical division [14].

Autopoiesis is a special case of homeostasis and it takes the position that metabolism and life is essential for grounding intentional categories like cognition, consciousness, and emotions [1]. In the second Volume of the Attachment Trilogy, Bowlby adopted the biological concept of homeostatis and applied it to behavioral as well as physiological control systems. In this view, physiological homeostasis which regulates food and sleep are an inner ring of control in the attachent control system. Attachment behavioural patterns constitutes an outer behavioral ring which is a complement to this inner physiological control system (Bowlby 1973, chapter 9). However, Bowlby did not set out how the intimate engagement of these two rings could give rise to phenomenoligical experience. He did describe attachment feelings, but within an emotional appraisal framework ([3], chapter 7). So viewing Attachment Theory through the lens of autopoietic enactivism can act as a spur for a more comprehensive approach that unifies behaviour, cognition, and subjective experience in a single explanatory framework.

4.3 A Radical Enactivist Manifesto for Attachment Theory?

Hutto and Myin propose the thesis of radical enactive cognition (REC) that is a variant of enactivism that states that only a small proportion of cognitive processing is mediated by contentful representations. In their view, the majority of human cognition is basic and non-contentful information processing that controls behaviour for adaptive purposes but does not possess truth bearing properties like reference, accuracy or implication. According to REC, contentful representations do mediate some cognition, but these representations play a minor role in cognition overall, *"emerging late in phylogeny and ontogeny, being dependent in special sorts of shared practices.* ([14], p 13). So what Hutto and Myin have proposed is a novel variant of a dual process approach to cognition, with linguistically mediated representations that can interpret or receive narrative meanings, and basic structures and mechanisms that carry out adaptive control [22]. However, whilst other dual process approaches make a distinction between self-reflective thought which is linguistically mediated

and conscious, and processing which is not linguistically mediated and inaccessible to consciousness, REC 'carves things up' in a very different way [22]. As Hutto and Myin note, *"Enactivists are concerned to defend the view that our most elementary ways of engaging with the world and others - including our basic forms of perception and perceptual experience - are mindful in the sense of being phenomenally charged and intentionally directed, despite being non-representational and content-free* ([14], p 13). So according to a REC approach to Attachment Theory, an IWM that is formed early in ontogeny and has become inaccessible to linguistic self-reflection is not 'hidden', or at 'behind' or 'beneath' other more linguistically accessible IWMs. Instead, REC reframes inaccessibility - so in REC this is just linguistic inaccessibility - so such inaccessible structures are still at the forefront of mind and are phenomenally charged and conscious. This reframing can turn therapeutic ideas right around. Instead of therapy uncovering hidden structures it is about understanding how context and behavioural predispositions enact these structures in the moments they occur.

In addition, REC holds that an organism's current behavioural tendencies are not explained or structured by representations of the past but influenced more directly, just by its *"history of active engagement."* with the world ([14], p 11-12). So an organism's behavioural predispositions do *"not inherently "say" anything about how things stand in the world"* ([14], p 19). Rather, according to Hutto and Myin, *"a truly radical enactivism - REC - holds that it is possible to explain a creature's capacity to perceive, keep track of, and act appropriately with respect to some object or property without positing internal structures that function to represent, refer to, or stand for the object or property in question"* ([14], p 82)

So if Attachment Theory follows REC it might reconceive internal states like working models to be just control states and break the link with the reality they are supposed to represent. An attachment control system that proposes internal control states are not truthful representations of reality is a profound shift from current Attachment Theory. No longer would attachment interventions be concerned to assess how individuals represented their past relationships but instead they would be more focused on how to move towards more adaptive behaviour patterns.

5 Conclusion

In breaking from psychoanalysis Bowlby was a revolutionary, but at heart he was also a conservative, because he wanted to save the core and most valuable findings of Freud's psychoanalytic framework. These were insights about the highly active and interactive nature of social and emotional development in infancy. Since Bowlby was an eager 'borrower' of scientific concepts from the ideas which were popular at the time he formulated Attachment Theory, he might today look to incorporate the diverse insights of 4e cognition in a revised framework for the attachment control system. In section 2 we asked whether IWMs in adults are linked both to processes of shared meaning making and interpretation, and to processes of adaptive control, that is, whether they should not only be conceived in linguistic or symbolic form, but also conceived as analogue or embodied information processing structures [24]. In section 3 we showed how extended cognition provides a possible explanation for how infants derive narrative meaning about their attachment relationships from their caregivers. Then in in section 4 we considered how an enactivist approach can help explain subjective experiences in attachment interactions, and how internal control structures can direct future actions without a link to 'truthful' representations of past events. Con-

sidering issues of embodiment, cognitive extension, and enactivism together has a major benefit because these three approaches pull in different directions. So together they provide a balanced reformulation. Considering IWMs as analogue in addition to symbolic keeps the IWM construct tied to an individual. The extended cognition approach reminds us of the dialogic nature of attachment and the enactive approach forces us to question our representational assumptions. Taken together these three perspectives complement each other. We can never really know how Bowlby would have responded to the questions posed by 4e cognition but we can act to make revisions to Attachment Theory that conserve his key theoretical insights.

REFERENCES

[1] M.A. Boden, *Mind as Machine: A History of Cognitive Science*, Oxford University Press, Oxford, 2006.

[2] J. Bowlby, 'The nature of a child's tie to his mother', *International Journal of Psychoanalysis*, **39**, 350–373, (1958).

[3] J. Bowlby, *Attachment and loss: volume 1 attachment*, Basic books, New York, 1969.

[4] J. Bowlby, *Attachment and loss: volume 1 attachment*, Basic books, New York, 1969 | 1982. (Second edition 1982).

[5] J. Bowlby, *Attachment and loss: volume 2, Separation: Anxiety and Anger*, Basic books, New York, 1973.

[6] J. Bowlby, *Attachment and loss: volume 3 loss, sadness and depression*, Basic books, New York, 1980.

[7] I. Bretherton and K.A. Munholland, 'Internal working models in attachment relationships', in Handbook of Attachment, (Second edition , *eds. J. Cassidy & P.R. Shaver*, 102–127, Guilford Press, London, (2008).

[8] W.J. Clancy, 'Scientific antecedents of situated cognition', in Cambridge Handbook of Situated Cognition*, eds. P. Robbins & M. Aydede*, 11–34, Cambridge University Press, New York, (2008).

[9] A. Clark, *Being There: Putting Brain, Body and World Together Again*, MIT Press, Boston, 1998.

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

[11] A. Clark and D. Chalmers, 'The extended mind', *Analysis*, **58**, 10–23, (1998).

[12] K. Craik, *The Nature of Explanation*, Cambridge University Press, London, New York, 1943.

[13] D.T. Gilbert, ' How mental systems believe', *American Psychologist*, **46**, 107–119, (1991).

[14] D. Hutto and M. Myin, *Radicalizing Enactivism: Basic Minds without Content*, MIT Press, Cambridge, MA, 2013.

[15] M. McGann, H. De Jaegher, and E. Di Paulo, 'Enaction and psychology', *Review of General Psychology*, **17**, 203–209, (2013).

[16] R. Menary, 'Introduction to the special issue on 4e cognition', *Phenomenological Cognitive Science*, **9**, 459–463, (2010).

[17] P. Paulconbridge, 'Hacking the Extended Mind', in *Proceedings of 'Re-conceptualizing Mental Illness: The View From Enactivist Philosophy and Cognitive Science', AISB Convention 2013*, 34–36, AISB Press, University of Sussex, Brighton, (2013).

[18] D. Petters, 'Simulating infant-carer relationship dynamics', in *Proc AAAI Spring Symposium 2004: Architectures for Modeling Emotion - Cross-Disciplinary Foundations*, number SS-04-02 in AAAI Technical reports, pp. 114–122, Menlo Park, CA, (2004).

[19] D. Petters, 'Building agents to understand infant attachment behaviour', in *Proceedings of Modelling Natural Action Selection*, eds., J.J. Bryson, T.J. Prescott, and A.K. Seth, 158–165, AISB Press, School of Science and Technology, University of Sussex, Brighton, (2005).

[20] D. Petters, 'Implementing a theory of attachment: A simulation of the strange situation with autonomous agents', in *Proceedings of the Seventh International Conference on Cognitive Modelling*, 226–231, Edizioni Golardiche, Trieste, (2006).

[21] D. Petters, ' Towards an Enactivist Approach to Social and Emotional Attachment.', in *ABSTRACTS. AISB50. The 50th annual convention of the AISB. Goldsmiths University of London.*, 70–71, AISB, Goldsmiths College, London, (2014).

[22] 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).

[23] D. Petters and E. Waters, 'Epistemic Actions in Attachment Relationships and the Origin of the Socially Extended Mind', in *Proceedings of 'Re-conceptualizing Mental Illness: The View From Enactivist Philosophy and Cognitive Science', AISB Convention 2013*, 17–23, AISB Press, University of Sussex, Brighton, (2013).

[24] D. Petters and E. Waters, 'From Internal Working Models to Embodied Working Models', in *Proceedings of 'Re-conceptualizing Mental Illness: Enactivist Philosophy and Cognitive Science - An Ongoing Debate', AISB Convention 2014*, AISB, Goldsmiths College, London, (2014).

[25] V. Reddy, D. Hay, L. Murray, and C. Trevarthen, 'Communication in infancy: Mutual regulation of affect and attention', in Infant development: recent advances*, eds. J.G. Bremner, A. Slater & G. Butterworth*, 247–273, Psychology Press, Hove, (1997).

[26] R. Rupert, 'Challenges to the hypothesis of extended cognition', *Journal of Philosophy*, **8**, 389–428, (2004).

[27] J. Small, *The Analogue Alternative: The Electronic Analogue Computer in Britain and the USA, 1930-1975*, Routledge, London, 2000.

[28] E. Thompson, *Mind in Life: Biology, Phenomenology, and the Sciences of Mind*, MIT Press,, Cambridge, Mass, 2007.

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

[30] R.A. Wilson and A. Clark, 'How to situate cognition: Letting nature take its course', in Cambridge Handbook of Situated Cognition*, eds. P. Robbins & M. Aydede*, 55–77, Cambridge University Press, New York, (2008).

[31] J.Z. Young, *A Model of the Brain*, Oxford University Press, London, 1943.