From Black Bile to the Bipolar Spectrum: A Historical Review of the Bipolar Affective Disorder Concept

The terms melancholia and mania have their etymologies in classical Greek. Melancholia is derived from ‘melas’ (black) and ‘chole’ (bile), highlighting the term’s origins in pre-Hippocratic humoral theories [1]. Where depression/melancholia was viewed as an excess of black bile, the humoral perspective saw mania as arising from an excess of yellow bile [2], or a mixture of excessive black and yellow bile [3]. The exact origins of the term mania however, are not as clear-cut as those outlined for melancholia. The Roman physician, Caelius Aurelianus, proposes several origins for the word mania, including the Greek word ‘ania’, meaning to produce great mental anguish. He also suggests ‘manos’, meaning relaxed or loose, which would approximate to an excessive relaxing of the mind or soul [4]. There are at least five other etymological candidates proposed by Aurelianus for the word mania and the confusion surrounding the exact etymology is attributed to its varied usage in the pre-Hippocratic poetry and mythologies [4].

The earliest existent conceptualisations of mania and depression (melancholia) as human ailments are found in the works of the Greek philosophers and physicians of the classical period. Primacy for the earliest systematic descriptions are typically attributed to Hippocrates (460 – 337 BC) [5]. It is important to note however, that in the classical conceptualisations of mania and melancholia the two entities are never explicitly related; there is no mention of an integrated manic-melancholic illness concept.

Arataeus of Cappadocia

The idea of a possible relationship between mania and melancholia is first alluded to in the 2nd century AD by Soranus of Ephedrus (98-177 AD). Soranus himself describes mania and melancholia as distinct diseases with separate aetiologies, however, he acknowledges that “many others consider melancholia a form of the disease of mania” [6]. For direct and unequivocal speculations about a relationship between mania and melancholia we must look to Arataeus of Cappadocia. Arataeus, an eclectic medical philosopher, living in Alexandria between 30 and 150 AD, is recognized as having authored most of the surviving texts referring to a unified concept of a manic-depressive illness [7]. Operating within the humoral paradigm, Arataeus viewed both melancholia and mania as having a common origin in excessive ‘black bile’ [7,8]. Furthermore, he connected the concepts, viewing them –mania and melancholia - as different aspects of the same illness with mania conceptualized as being an end stage of melancholia [4,5,8]. In his “On Etiology and Symptomatology of Chronic Illnesses”, Arataeus writes:

“…I think that melancholia is the beginning and a part of mania... The development of a mania is really a worsening of the disease [Melancholia] rather than a change into another disease...In most of them [melancholics] the sadness became better after various lengths of time and changed into happiness; the patients then developed mania.” [4].

And also: “…the melancholia increases and becomes mania.” [8]

The writings of Arataeus represent the earliest known written record of the birth of the bipolar disorder / manic depressive illness concept. His ideas resonate with the manic-depressive illness concepts of more recent centuries. For example, Arataeus’ description of the transformation of depression into mania aligns with the manic defence hypothesis. This is an idea detailed by psychoanalytic writers, contending that mania is an extreme defence against, or reaction to, depression (Abraham, 1911/1927; Dooley, 1921; Rado, 1928). This is an idea that has, in more recent years, received empirical support through the work of contemporary experimental psychologists (Bentall & Thompson, 1990; Neale, 1988; Thomas & Bentall, 2002; Winters & Neale, 1985).

Arataeus’ bipolar concept also included the idea of predisposing traits, suggesting individuals who developed mania were characteristically labile, irritable, angry or happy. Conversely, those who only developed melancholia were viewed as tending towards depression in their pre-morbid state (Zax & Cowen, 1976). Arataeus ultimately saw emotional disorders as magnifications or exaggerations of existing character traits, another idea that would be further explored in later centuries with reference to the idea of fundamental states, i.e. cyclothymic, hyperthymic and dysthymic (non dihethmic) temperaments [7]. Arataeus also observed that both mania and depression could occur simultaneously in the same person, an observation according to early 20th century nosological concepts such as ‘mixed states’ (Kraepelin, 1976) as well as contemporary diagnostic categories and bipolar spectrum concepts such as depressive mixed states (Benazzi 2005). Despite the many striking conceptual comparisons, Arataeus’ notions of mania and melancholia were much broader than our current nosologies, arguably including schizophrenia, schizoaffective disorders, psychotic depression and organic psychoses [4].
After Arataeus

Over the intervening centuries many physicians have shared Arataeus’ ideas about a close connection between mania and melancholia. Examples include Paul of Aegina (625-690) and Paracelsus (1493-1541). Notably, in 1567 Alexander of Tralles, a Byzantine physician, acknowledged that cases of chronic melancholia could be associated with recurrent, or periodic attacks of mania in a cyclical manner. Alexander also goes on to suggest that the features of mania and melancholia often intermix within a single episode/attack [9]. Other physicians who maintained the manic–melancholic association were Thomas Willis (1621–1675) and Giovanni Morgagni (1628–1771). There were those however, who saw things differently. Taking the opposing view are Timothie Bright (1550–1615), Robert Burton and Phillippe Pinel (1745–1826).

Folie circulaire

This debate about the relationship between mania and depressive states was finally resolved in the 19th century. Specifically, the contemporary psychiatric conceptualisation of manic-depressive illness is typically traced back to the 1850s. Marneros [8], describes the concepts emerging out of this period as the “rebirth of bipolarity in the modern era”. This rebirth began on January 31st 1854, when Jules Baillarger described to the French Imperial Academy of Medicine a biphasic mental illness causing recurrent oscillations between mania and depression. To slightly complicate issues, two weeks later, on the 14th February 1854, Jean-Pierre Falret presented a description to the Academy on what was essentially the same disorder. This illness was designated folie circulaire (‘circular insanity’) by Falret, and folie à double forme (‘dual-form insanity’) by Baillarger [10]. Baillarger made accusations of plagiarism and contested the issue of precedence. To vindicate himself of these accusations, Falret produced a 14-sentence long report on his folie circulaire that he had had published in the Gazette des Hopitaux several years earlier, in 1851. Baillarger, unconvinced, continued to reiterate and extend these accusations of plagiarism until his death in 1890 [11].

Falret [12], went on to publish the more substantial description of folie circulaire: “Mémoire sur la folie circulaire, forme de maladie mentale caractérisée par la reproduction successive et régulière de l’état maniaque, de l’état mélancholique, et d’un intervalle lucide plus ou moins prolongé”, an approximate English translation being; “Circular insanity (is) a form of mental disease characterized by the successive and uniform reproduction of the manic state, melancholic state, and of a lucid interval of varying duration”.

On the issue of priority, Pichot argues that an objective analysis of the printed material in relation to the two concepts demonstrates that Baillarger’s accusations were unfounded. He suggests that the two concepts, folie circulaire and folie à double forme, differ on many important points, with Falret's being closer to our present conceptualisation. Many other commentators share this view [4,5,9,10]. Baillarger’s folie à double forme concept assumed a disease in which depression and mania change into one and other, but the interval between transformations has no importance. Falret, however, included a longitudinal perspective, which included the possibility of a life-long disorder. Falret also considered the pattern of phases of mania and melancholia including the inter-phasic interval (euthymia), thereby allowing episodes of mania and depression separated by a long interval to be still considered folie circulaire [9].

Berrios [13], argues that the primacy debate is futile, pointing out that there were several other French authors - contemporaries of Baillarger and Falret - also writing along the same lines: Billod (‘folie à double phase’, ‘dual-phase insanity’) and Legrand du Saulle (‘folie alteme’, ‘alternating insanity’). However, despite these emerging 19th century conceptualisations of a unitary biphasic disorder and their spread across Europe and the US [8], there were still many clinicians who continued to view mania and melancholia as invariably discrete, separate entities. It took the pioneering work of Emil Kraepelin to firmly establish the concept of manic-depressive insanity [5].

Manic depressive insanity

Emil Kraepelin was born in Neustrelitz Germany 1856. Graduating in medical studies in 1878, he gravitated towards further study and work in the field of psychiatry, initially travelling to Munich to work with neurobiologist Bernhard Von Gudden; however, his poor eyesight made microscope work difficult [6], and in 1882 Kraepelin left Munich and came to study under Flüssig in Leipzig. This did not work out and eventually Kraepelin was taken under the wing of Wilhelm Wundt, working in Wundt’s psychological laboratory in Leipzig. Wundt encouraged Kraepelin to write his ‘Compendium of Psychiatry’, a publication that would eventually have a near-revolutionary impact on the field of psychiatry [14].

By the second edition of the ‘Compendium’ (1887), Kraepelin was suggesting that mental illnesses could be identified and organised into a small, discrete number of categories, initially identifiable by symptomatology. Kraepelin collected hundreds of case studies and concluded that symptom groups followed characteristically different courses, eventually arriving at three categories: dementia praecox, paranoia and manic depressive insanity.

By the sixth edition (1889), the term ‘manic depressive insanity’ (manisch-depressives irresein) had been born [6]. Kraepelin’s manic depressive insanity evolved into a broad category that eventually encompassed virtually all forms of melancholia and mania, including what would today be considered unipolar depression. Kraepelin essentially put folie circulaire and melancholia together to create his unitary illness entity [8]. Kraepelin reasoned from clinical experience that if a whole series of manic episodes could unexpectedly be punctuated by a depressive episode, and thereby be considered circular insanity, then this possibility should be extended to ‘periodic melancholia’.

There were several key additions to the eighth edition of the compendium specifically the idea of mixed and fundamental states. Mixed states represent symptom mixtures, e.g. depression with flight of ideas. Fundamental states are characterized as temperaments or dispositions corresponding to less severe and more enduring versions of the morbid affective states. Both the concepts of mixed states and affective temperaments are central to what contemporary theoreticians describe as the bipolar spectrum.

As influential, pervasive and durable as it has been, Kraepelin’s
system of classification was also initially the target of the great deal of criticism, much of which still reverberates within contemporary nosological debates. Hoche [15] directed his critique at the system’s assumption of a linear relationship between clinical symptoms and localized brain lesions or micro-chemical alterations. Hoche argues that essential psychotic symptomatology involved such things as affect, will, and judgment, all of which engage widely distributed brain areas, and therefore trying to map mental disease entities to anatomical changes, would be futile [16]. Bonhoeffer [17] held the view that aetiological differentiation was only possible at the somatic or neurological level, but not at the level of psychological symptomatology. Bonhoeffer uses the analogy with alcoholism, illustrating how the same aetiology can give rise to varying disease entities, e.g. delirium and hallucinosis, but also stressing the converse, that is, that diverse aetiological factors can give rise to identical clinical manifestations [16].

Kraepelin’s broad manic-depressive illness group also met with some opposition in relation to its all-inclusiveness, for example in Scandinavia, Lange, Christiansen, Pendersen and others continued to work with periodic depression as a separate affective disorder [18-20]. Similarly, Bennon [21], argued for separating periodic depression from manic-depressive illness; his call, however, was met with little approval [4]. In Germany, too, Carl Wernicke challenged the Kraepelinian view of manic-depressive illness inclusive of unipolar depression and mania. Wernicke contended that illnesses where there are only recurrences of depression or only recurrences of mania are distinct from manic-depressive illness [4]. Likewise, Adolf Meyer viewed the collapse of the various manifestations of mood disorders into one group as a ‘startling condensation’ [22]. Of this monolithic category Karl Jaspers writes:

“from time to time in psychiatry, there emerge diseases which constantly enlarge themselves until they perish from their own magnitude.” [23,6]

Despite some initial challenges, Kraepelin’s viewpoints and his ultimately dichotomous taxonomy of psychotic illness gained wide acceptance in a relatively short period of time, contributing to a relative conceptual unification of European psychiatry [5]. Bentall [21] suggests that,

“This ultimate triumph partly reflected the simplifying effect that Kraepelin’s ideas had on the theory and practice of psychiatry.”

This observation is also reflected in Goodwin and Jamison’s [5], attitude towards categorical, as opposed to dimensional approaches to conceptualisation with categorical approaches viewed as intrinsically easier to understand and manage statistically.

Post-Kraepelinian concepts of manic-depressive illness evolved differently in Europe and the USA. European psychiatrists maintained fidelity to a more traditionally-rooted medical disease model of mental illness, whilst their North American counterparts were greatly influenced by psychoanalytic perspectives, and came to place an increasing emphasis on psychosocial factors in their understanding of mental illness [5].

Adolf Meyer was instrumental in reshaping the conceptual framework adopted in North America [5]. The emergent Meyerian framework viewed biological and genetic factors as underlying vulnerabilities to specific psychosocial influences. This conceptual shift is illustrated by the 1952 American Psychiatric Association diagnostic manual’s (DSM-I) description not of manic-depressive illness, but rather of manic-depressive reaction, conceptualised as a subcategory of affective reactions [5].

In Europe, the work of Eugene Bleuler would extend, and to some extent challenge, the Kraepelinian legacy. Bleuler, like Kraepelin, drew on his observations of patient’s symptoms. Bleuler [24], came to view Kraepelin’s term ‘dementia praecox’ (precocious dementia) as misleading, given that the illness’ onset was not exclusively associated with adolescence, nor did it invariably result in extreme mental deterioration. Bleuler adopted the term schizophrenia, which better describes what Bleuler saw as the core of the illness, specifically, a separation between the functions of personality, thinking, memory and perception [21]. Bleuler essentially broadened the Kraepelinian boundaries in relation to schizophrenia/dementia praecox, since his concept had room for a ‘simple schizophrenia’ (a form of schizophrenia without delusions or hallucinations) and the seemingly subclinical “latent schizophrenia” which cast the shadow of potential case-ness over “… irritable, odd, moody, withdrawn or exaggeratedly punctual people…” [24,21].

In 1924, Bleuler’s analysis of the ‘psychoses’ focused on the relationship between the Kraepelinian conceptualisation of manic-depressive illness and Bleuler’s broader conceptualization of schizophrenia. Ultimately, Bleuler came to view the demarcation between these two categories of illness as wholly superficial and he proposed a continuum between the two. For Bleuler, an individual could be at different points along this continuum over the course of their illness [21]. Kraepelin himself, in 1919, acknowledged that features of the two illnesses were at times indistinguishable [25]. Bleuler’s concept placed manic-depressive illness on a continuum with schizophrenia, the exact distance between the two constructs being the degree of schizophrenic symptomatology. Bleuler identified four “fundamental symptoms” of schizophrenia, sometimes known in the English-language speaking world as the Bleuler’s four ‘A’s (loosening of Associations, Ambivalence, Autism and inappropriate Affect) [21]. For Bleuler, the presence of any of these symptoms required a diagnosis of schizophrenia. Bleuler considered affective symptomatology as non-specific, a diagnosis of manic depression being made only after the exclusion of schizophrenia [21], concordant with the dictum “even a trace of schizophrenia is schizophrenia” [26].

Bleuler also broadened the manic-depressive illness category to include several subcategories and adopted the term ‘affective illness’. Goodwin and Jamison [5], suggest that the influence of Bleuler’s re-categorisation and conceptual broadening are evident in the International Classifications of Diseases (8th and 9th editions) and the Diagnostic and Statistical Manuals of the American Psychiatric Association (DSM I and II), for instance, in the inclusion of schizoaffective illness as a subtype of schizophrenia. They also remark that Bleuler’s addition of subcategories of affective illnesses anticipated the contemporary unipolar–bipolar subdivision of the Manic-Depressive diagnostic group.
The eventual subdivision of the Manic-Depressive diagnostic group was also greatly influenced by the work of Leonhard, who in 1957 proposed a classification system that made a distinction between patients with a history of depression and mania and those with a history of only depressive episodes [5]. Leonhard’s work had its roots in the work of Wenicke who, as previously mentioned, opposed Kraepelin’s inclusion of pure phasic mania and depression in the manic-depressive illness construct. Taking this idea forward, Leonhard termed those with a mixed history (episodes of mania and depression) ‘bipolar’, and those with a history of depression or mania only ‘monopolar’ [5]. Leonhard also noted that his bipolar patients had a higher incidence of mania within their families compared to the monopolar patients. This distinction was subsequently substantiated by family history data [27–29]. These findings suggest that the most prevalent affective disorder in the relatives of bipolar patients is unipolar disorder followed by bipolar disorder, the rates of which are two to three times greater than the rates of affective disorders in the relatives of control patients. The rate of bipolar disorder in the probands of unipolar patients is only marginally, and not always significantly, higher than the rate found in the control group’s probands [5].

Angst (1966/1973) [27], offered further data supportive of a bipolar–unipolar differentiation. Angst studied 326 patients at the university hospital in Zurich between 1959 and 1963 and drew attention to several differences between unipolar and bipolar disorders; for example, he noted that for bipolar disorders the gender ratio was fairly equal, whereas for unipolar disorders it was elevated in females. Similarly, late onset depression was associated with unipolar, but not bipolar disorders. With reference to Angst’s (1966) publication, Pichot [30], asserts that the concept of bipolar disorder was reborn, the first birth being attributed to fellow Frenchman Faltet, and surprisingly not Arataeus of Cappadocia.

Contemporary conceptualizations

Manic-depressive “illness” is currently conceived of as a recurrent biphasic affective/mood disorder including episodes of hypomania, mania and depression. Manic-depressive illness has dropped the arguably metaphoric “illness” appendage, [31] and to distinguish it from unipolar depression is today termed bipolar disorder or bipolar affective disorder. Unlike dementia praecox, however, this disorder appears to be fairly resilient to re-branding, and manic depression is still a term in common use by the general population as well as many professionals writing about the illness/disorder. Many of the large self-help groups for people who have experienced bipolar disorder still retain the name ‘manic depression’; the Manic-Depression Fellowship, for instance, only recently chose to re-brand itself as “MDF The Bipolar Organisation”. More than a quarter of a century after the “illness” was recategorised and renamed bipolar disorder in DSM-III [32], it is still commonly introduced to the public as ‘bipolar disorder, formally known as manic-depressive illness’. Goodwin and Jamison propose that manic-depressive illness represents the “... magnification of common human experience” [5] and also suggest that “few maladies in medical history have been represented by such unvarying language” (1990 p.56) [5].

It is this magnification of common human emotional experiences and the related descriptors ‘mania’ and ‘depression’ that arguably give the classical/Kraepelinian nomenclature ‘manic-depression’ its durability. In contrast, the term ‘bipolar’ is descriptive of the disorders’ course but silent about the emotional content - the common human experience.

Affectophilia or the bipolar spectrum?

Akiskal et al. [33], suggest that “…the uncoupling of depressive disorders from the more strictly defined bipolar disorders...left undefined many affective conditions lying in the interface of unipolar and bipolar disorders.” (S7).

Akiskal et al. [33], contends that the post-Kraepelinian conceptualisation of manic depressive illness has resulted in a too narrowly defined bipolar category, and an implausibly broad heterogeneous group of conditions labelled major depression [34]. Of particular consequence is the definition of hypomania; for Akiskal et al. [33], the duration threshold of 4 days is too high. Cassano et al. [35], used a two-day duration threshold for hypomania, as opposed to the four days proposed within the DSM-IV. They found that patients diagnosed according to the two-day duration had rates of familial bipolarity statistically indistinguishable from bipolar I patients, both of which were significantly higher than the rates found in unipolar major depressive disorder patients. Angst and Gamma [36], propose dropping the duration criteria altogether, with reference to the occurrence of very brief episodes of hypomania observed in adolescents with bipolar disorder.

Even leaving aside the duration threshold issue, hypomania, and therefore bipolar II, are still particularly difficult to diagnose. The under-diagnosis of hypomania, some argue, has important clinical implications. Angst and Gamma [36], suggest that any failure to diagnose hypomania leads to a false positive diagnosis of major depressive disorder (MDD). Angst and Gamma maintain that this is the case in up to 50% of MDD diagnoses.

In terms of improving the diagnostic criteria for hypomania, both Akiskal et al. [33] and Angst and Gamma [36], argue for the central importance of over-activity as obligatory criterion in the diagnosis of hypomania, with Akiskal et al. placing it above changes in mood in terms of its diagnostic importance. Angst and Gamma [36], more conservatively suggest that either mood change (euphoria/ irritability) or over-activity should represent the obligatory criteria for hypomania. A further difficulty, not addressed by lowering duration thresholds or changing criteria, is the fact that hypomanic episodes may not be particularly distressing for the individual and may be experienced as ‘normal functioning’, ‘extreme wellness’, ‘intense creative episodes’ etc. [37]. For this reason, detection may often have to be retrospective, if not reconstructive, and reliant on third party reports of the affective states and behaviours. This presents a situation where identifying hypomania is reliant on the vagaries of memory and/or whether or not significant others/carers etc. are available for interview.

For Akiskal the splitting of bipolar and unipolar depression, coupled with the conceptual and operational difficulties surrounding hypomania, has left a group of disorders that are (a) ill-defined, (b) undefined and (c) difficult to diagnose [7]. Cassano, et al. [35], share...
this view; the title of their special article summed up the situation as “The bipolar spectrum: a clinical reality in search of diagnostic criteria and an assessment methodology”. This view that much bipolar illness goes unrecognised, perhaps buried within the ‘impossibly broad major depressive disorder (MDD) group’, gives rise to questions of quantification. Akiskal, et al. [33], assert that between 30 and 55% of all affective disorders are found within the broader bipolar spectrum, including many undefined ‘subthreshold’ expressions. As previously mentioned, Angst and Gamma project that 50% of MDD cases are actually undetected bipolar II. For Akiskal there is also bipolar III, and IV.

These undefined ‘subthreshold’ conditions are reported to have adverse psychosocial consequences [7]. Such consequences are arguably exacerbated by a lack of diagnostic and assessment methodology [35], leading to delayed and under-diagnosis [38,39].

Is the future of bipolar disorder about greater diagnostic granularity, a bipolar spectrum if you will? A more viable alternative would be to take a symptom-centric approach [14]. This would involve greater focus on the specific complaints associated with each case: for example over activity (reward sensitivity), grandiosity, insomnia etc. rather than the specific diagnosis assigned. In terms of research, at example over activity (reward sensitivity), grandiosity, insomnia etc. rather than the specific diagnosis assigned. In terms of research, at

Acknowledgement

We would like to thank the Al Jalla Foundation for supporting this work.

References


