



## Biopsychosocial Conditions of Health and Disease

**Abstract** This chapter continues from the previous chapter on themes in biopsychosocial conditions of health and disease, picking up some core questions familiar in the theory and philosophy of medicine. We argue that the concepts and boundaries of health and disease are themselves biopsychosocial. Controversies about whether such-and-such a condition is or is not a medical matter, as opposed to difference or lifestyle choice, the consequences of being which involve benefits such as access to healthcare and/or harms such as stigma, and the terms in which such debates are conducted—are all thoroughly biopsychosocial-political. Core defining features of illness—activity limitations, pain and distress—likewise involve our psychology and social life as well as our biology. On the theme of causation, we endorse scientific method as the route to identifying causal mechanisms, note the major role of chronic stress in models of causal mechanisms linking psychosocial factors with biological damage, and spell out that chronic stress is a quintessential biopsychosocial concept. We consider the Research Domain Criteria (RDoC) proposed recently by the N.I.M.H. as a framework for research in mental health as an illustration of a biopsychosocial research framework, potentially extendable to cover physical health and biomedicine. Physical and mental health conditions are brought together in the new biopsychosocial model rather than being axiomatically separate—as they were in the old context of reductionism and dualism.

**Keywords** Biopsychosocial research framework · Causal mechanisms · Patient-centred care model · Recovery model · Pain · Social disability model · Research Domains Criteria (RDoC)

#### 4.1 CONDITIONS OF BIOPSYCHOSOCIAL LIFE

So far, we have reviewed the rationale as well as the challenges for the biopsychosocial model, in Chapter 1, and, drawing on contemporary life and human sciences, presented conceptualisations of the biological, in Chapter 2, and the psychological and social, in Chapter 3. In the later parts of Chapter 3, we drew out features of the biopsychosocial whole, especially that causal interactions run within and between these three, or rather four, since the natural environment is thoroughly involved with all, a point made explicit in Sect. 2.3. Critical for the biopsychosocial model as a model of health and disease, foundational biopsychosocial concepts already have the relevant normative distinction built into them. Chapter 2 characterised biological processes as local areas running contrary to the general direction of the second law of thermodynamics, temporarily; this feat achieved by metabolic regulatory mechanisms, which, as and when they fail, jeopardise the viability of the organism, which in any case eventually inevitably ends up back as dust, more or less prematurely. Some familiar conditions of biological life were listed in Sect. 3.3, such as food, water, oxygen, suitable ambient temperature, accommodation, on the way to the less obvious conditions of psychological life as agency, which include opportunity, access to resources such as education and training, which involves, picking up themes in Sect. 3.2, recognition by a supportive, encouraging and resourced social group, which is also necessary, along with peace and an uncompromised natural environment, for basic biological security.

In short, the conditions of biological, psychological and social life—risks and protective factors for developing and recovering from health conditions—are pervasive. As this has become clear in science and our thinking, health policy has recognised that it should be cross-sectoral, involving much more than the healthcare sector. As a general rule, the healthcare sector is trained and funded to treat people who have fallen below the or their normal level of functioning, in important domains, with the aim of restoration of function, or optimising lowered function,

or reducing avoidable further loss. The healthcare sector also has a limited role in primary prevention, but this has turned out to be a multifaceted, cross-sectoral task, in which the illness-related language of healthcare gives way to more general concepts such as promoting resilience, or well-being, thriving, or further away still, happy enough family life, access to and use of educational opportunity, satisfying and well enough paid work, friendships and meaningful civic engagement.

The biopsychosocial model is a model of health and disease, but this roughly divides up, for reasons understandable in terms of the model, into, on the one hand, the business of the healthcare sector—illness, with the negative conception of health as avoidance of or recovery from illness—and, on the other hand, prevention of illness, which merges into resilience and thriving, which are protections against ill health from the point of view of healthcare, but which from all other, non-illness preoccupied sectors, are another thing altogether: education, work, economics, politics, environmental policy and security. In this fourth chapter, we pursue further implications of biopsychosocial theory as a model of disease and health as absence of disease, leaving aside the broader questions of health and well-being.

## 4.2 BIOPSYCHOSOCIAL CONCEPTUALISATION OF HEALTH CONDITIONS

### *Concepts and Boundary Disputes*

Some questions about illness are causal and some conceptual, though there is no hard and fast line between them. Causal questions are about identifying risks, pathways and mechanisms of disease, developing clinical therapeutics and treatment technologies. They are questions for the basic and clinical sciences, and they take for granted what is disease and what is not. Or rather, to make clearer the conceptual point, the sciences are independent of whether the conditions of interest are called ‘disease’, and who by, as opposed to, for example, just phenotypic ‘difference’ or ‘life-style choice’, or ‘sin’. These alternatives show up not much in science, but more in social, political, legal, moral and theological debate. They raise more or less far-reaching conceptual issues, to do with the meaning of ‘illness’, the logic of illness attribution and human nature. While scientific findings may be brought to bear on them, typically they cannot be settled by the science, or at least not to everyone’s satisfaction,

indicating that there are other relevant considerations. The main point of this section is that these conceptual questions are biopsychosocial. Here is Karl Jaspers on this point, in 1913 around the beginning of modern medicine and psychiatry, also noting that such questions do not normally merit medical attention [1] (p. 652):

What health and illness mean in general are matters which concern the physician least of all. He deals scientifically with life processes and with particular illnesses. What is 'ill' in general depends less on the judgement of the doctor than on the judgment of the patient and on the dominant views in any given cultural circle.

In conditions of certainty, there is no need to spell out what *illness* means; all concerned, at home and in the hospital, know only too well, and they have other serious tasks to be getting on with. And if and when the need for a definition of *illness* or *disease* arises, in a textbook or classroom or the clinic, it is easy enough to give one simply by using some other term with similar meaning, such as *abnormal* structure or function, or *disruption*, *disturbance*, *dysregulation*, etc. The medical textbook descriptions of signs, syndromes and diseases are full of such terms, and they can all be used to define each other well enough for most purposes. This family of disease/illness related terms are all typically systemic, referring to abnormalities or disturbances of structural forms, regulation and functional ends.

The meaning of illness—and all the cognates to do with abnormality—becomes an issue in conditions of uncertainty and dispute as to whether such and such a condition is an illness or not, in circumstances when it seems that no further observation or laboratory test would settle the matter clearly one way or the other. This uncertainty arises when criteria that normally go together, in the paradigm or prototypical kind of case, fall apart. Three key features of illness typically go together: the person complains of distress or pain; second, they are unable to do things they need to do, there is incapacity or activity limitation, loss of agency; and third, there is the assumption that these things are because something is not well with the person's body or mind. This last assumption implies that medical/psychological expertise is required, hopefully, to reduce the harm (the distress/pain and incapacity) and not create more. When these features and assumptions are all present and correct, there are conditions for certainty—but insofar as they cleave apart, some

present, some absent, or dubious, the position becomes ambiguous; attribution of illness and the closely linked perceived need for healthcare professional attention, become uncertain. With uncertainty comes controversy. Examples of general kinds of case where attribution of illness and/or need for medical attention is commonly contested include: some mental health diagnoses, especially those associated with non-voluntary admissions; ‘medically unexplained’ conditions; alleged over-medicalising and overtreatment of conditions that are regarded rather as ‘normal’, self-limiting, or less harmful compared with harms from treating; pathologising/stigmatising difference and diversity, and lifestyle choices—even if they carry raised risk of illness.

The reference to ‘pathologizing/stigmatising’ refers to a downside of the illness attribution—a harmful side effect, linked to the main effect. The main effect of illness is to decrease our agency, up to and including the ultimate ending of it all. The adverse effects of illness on agency, and the experience of illness as pain and distress, cue interpersonal and institutionalised responses to help, to provide resources including care and treatment. At the same time, social expectations are reduced: the ill person is excused from normal social role obligations, from moral responsibility, and hence from blame; the attribution of causation of behaviour, or of inactivity, is to the illness, not to the person as agent. In short, illness attribution implies excuse for and not being blamed for the downturn in functioning, and the right to access available health-care. However, at the same time, and for the same reasons, pathologising carries the risk of disqualification of the person from their full recognition as a self-determining agent in the social world—with knock-on risks of being stigmatised and subjected to the many varieties of social exclusion. The relative benefits and losses of being seen as ill depend on many general and individual factors. For sudden onset, treatable conditions, the benefits are typically high and the costs (of the above psychosocial sort) relatively low. For long-term conditions, with no treatment, or with treatments that—from the person’s point of view—may do as much harm as good, the balance shifts, costs may outweigh benefits, leading to a rejection of the ‘illness’ label, in favour of ‘difference’ and ‘diversity’.

Conditions of so-called ‘disability’ are a special case, distinguished from ‘illness’, for complex reasons including absence of treatment, and, commonly, the absence of complaints of distress. However, the ‘disability’ label functions like the ‘illness’ label in that it imputes deficit in

ability, hence excuse from ‘normal’ social role obligations, and hence at the same time carries risk of demotion from full recognition as an equal agent. Furthermore, for lifelong conditions, ‘disability’ is not relative to the person’s own previous functioning, and for acquired chronic conditions, the downturn in personal functioning does not come with implied upturn, and becomes ‘difference’. The implication is that the critical notion of ‘deficit in relation to normal’ is benchmarked not against the person’s own normality but against normality of the majority. Further, since ability to act is a function not only of personal abilities but that always in the context of task demands and available resources and opportunities, the cause of ‘disability’ can be legitimately attributed to these external factors; for example, people who have to use wheelchairs are handicapped in mobility by the way the majority build transport systems, not by their condition in itself. These kinds of points are well theorised in the ‘social model of disability’ (e.g. [2, 3]).

In short, conceptual issues around illness—and disability-related concepts and practices involve a complex range of and interaction between biological, psychological, social, moral and policy factors. They cannot be so much as articulated without a full biopsychosocial framework.

### *The Logic of Disease Attribution Is Top-Down*

The fundamental feature of disease in a system is that it causes—or significantly raises risk of—disruption of the function of the system, thus leading to adverse outcomes for a dependent system. ‘Adverse outcomes for a dependent system’ implies disruption of functionality of that dependent system, and that dysfunctionality in turn means that it causes or raises of adverse outcomes on a further dependent system... and so on. This cascade continues until we reach dysfunctionality/adverse outcomes for the organism as a whole in its activity in the environment. It is poor outcomes at the level of the whole that ultimately drives attribution of dysfunctionality downwards to the parts that serve the whole. In short, the difference between function and dysfunction (or between good enough function and not good enough function) of parts, ultimately turns on the difference between function and dysfunction (or between good enough function and not good enough function) of the organism as a whole.

In short, the logic of disease attribution is top-down, not bottom-up. The causal pathways, by contrast, can be bottom-up, as for example

atherosclerosis causing an embolus causing cardiac arrest causing the person's death. Causal pathways can also be top-down as previously considered; for example, chronic unhealthy diet and lack of exercise raising risk for atherosclerosis. But the point made here is a different one, namely, that the logic of dysfunction, as opposed to function, is top-down, in the sense that it flows from whole to part.

### *The Centrality of the Person*

The implication of the line of argument in the preceding section is that, in biopsychosocial systems theory, health conditions have to be understood in terms of the person as a whole, specifically how it affects their agency, values and achievement of personal goals. The centrality of the person was identified by Engel as a feature of the biopsychosocial model in his original papers [4, 5]. In this respect, there are some connections between the biopsychosocial model and the Patient- or Client-centred Care Model, and many papers have examined the relation between the two approaches (e.g. [6–9]).

There are also connections with the Recovery model, a relatively recent, important and radical influence on mental health services [10–12]. The model focuses on chronic health conditions, on the centrality of the person's life and values, on achieving a good quality-of-life, on the need for good medical and nursing care, especially in acute phases, and on issues of access to social resources and opportunities. Many if not all aspects of the Recovery model can be applied as much to chronic physical health conditions as to mental health conditions, though so far this extension is in its infancy (see e.g. [13]).

The focus on chronic conditions has implications for a distinction often drawn in the literature between *illness* and *disease*, illness being the condition of the person, and hence involved in the personal and social world, and disease being a dysfunctional condition of a bodily organ or system, and not so involved (e.g. [14, 15]). Although this is an important distinction, it applies most clearly in cases where there is an identifiable somatic disease process, in an otherwise unaffected person, which subsequently remits, spontaneously or with treatment. The distinction becomes blurred, however, in chronic conditions that have to be accommodated in the person's life and hence involved with the development of the whole personality. William Osler's famous remark: It is much more important to know what sort of a patient has a disease than what sort

of a disease a patient has [16]—applies especially well to chronic conditions. This blurring of the difference between disease and illness is another aspect of the shift from the (infectious) disease paradigm to focus on non-communicable conditions, which require a biopsychosocial approach.

This shift affects the definitions of health and disease. A recent BMJ article by an international group [17] critiques the 1948 WHO definition of health as ‘complete physical, mental and social well-being and not merely the absence of disease or infirmity’. The authors acknowledge the definition’s groundbreaking breath and ambition at the time, going on to criticisms, mainly unintentional contribution to the medicalisation of society, continuing to the second problem, arising from the increasing relative prevalence of long-term conditions [17] (p. 1):

The number of people living with chronic diseases for decades is increasing worldwide... In this context, the WHO definition becomes counterproductive as it declares people with chronic diseases and disabilities definitively ill. It minimises the role of the human capacity to cope autonomously with life’s ever changing physical, emotional, and social challenges and to function with fulfilment and a feeling of wellbeing with a chronic disease or disability.

Consistent with this last sentence, the authors go on to propose a conceptualisation of health as the *ability to adapt and to self-manage*, discussing this in relation to physical health, mental health and social health.

A concept of health along these lines makes personal agency fundamental, though in its broad biopsychosocial context, interacting with resources and opportunities. A person’s sense of agency, whether they can do enough to have a viable life, and with it whether they wish for life, as it is, or better, with treatment or with none, are all matters that depend on the person. The condition of the biological body matters, but insofar as it affects the person. It is of fundamental importance in health-care that it is the person who feels ill and wants treatment, any or more, or feels well enough without it. We attend to the person, not the body part—and not to psychological signs and symptoms in isolation either. The centrality of the person also shows up in the next section. We will quote from Wittgenstein: the pain may be in the hand, but we comfort the person, not the hand.



### *Pain and Distress Have Personal Biopsychosocial Meaning*

Activity limitation is the core behavioural feature of illness or injury, *pain* and *distress* are their subjectively experienced aspects. But even these subjective experiences turn out to be thoroughly biopsychosocial, whichever way one approaches them: by philosophical analysis of ‘subjective experience’, or in terms of neuropsychological models of causal pathways, or behavioural models of interpersonal pain signalling functions.

Cartesian dualism has had a massive impact on our folk way of thinking about subjective experience, especially inclination to suppose that it is essentially *separated from the body* and *private to the person*. One of the key philosophical critiques of what could be called ‘folk dualism’ (as opposed to the full-blown complex, original Cartesian metaphysics linked to the mechanisation of the world-picture) is by Ludwig Wittgenstein in his *Philosophical Investigations*, much of which anticipates the kind of philosophical framework we are proposing in this book. Here is a conclusion of Wittgenstein’s discussion of the meaning of ‘pain’ [18] (para. 281):

‘But doesn’t what you say come to this: that there is no pain, for example, without *pain-behaviour*? It comes to this: only of a living human being and what resembles (behaves like) a living human being can one say: it has sensations; it sees; is blind; hears; is deaf; is conscious or unconscious.

Expressed in the terms of ‘dynamical forms’, which we have used throughout in explicating biopsychosocial theory, it is the human living form that has psychology. This is a fundamental aspect of the non-dualist idea of ‘embodiment’, considered previously in Sect. 3.1. In the same movement of thought, Wittgenstein finds a simple way of making clear that the human being has a special ‘centre’—the ‘I’, the person’s speech and face—involved in recognition [18] (para. 286):

But isn’t it absurd to say of a *body* that it has pain?—And why does one feel an absurdity in that? In what sense is it true that my hand does not feel pain, but I in my hand? What sort of issue is: Is it the *body* that feels pain?—How is it to be decided? What makes it plausible to say that it is *not* the body?— Well, something like this: if someone has a pain in his hand, then the hand does not say so (unless it writes it) and one does not comfort the hand, but the sufferer: one looks into his face.

Reconstructions of pain that involve moving away from preconceptions of mind/body dualism, appear increasingly in the humanities literatures (e.g. [19, 20]).

Turning to the science of pain, this is an expanding, large and complex area of research and we give a very brief and simplified survey of some relevant key points. Until about the mid-1960s, it was supposed that pain was caused by signalling of tissue damage to the central nervous system by specific pathways. In the 1960s, Ronald Melzack and Patrick Wall [21] proposed their innovative ‘gate theory’ to supersede specificity models, explaining how pain perception involved multiple neural pathways, creating a model more able to capture individual and cultural contextual factors in pain perception. Melzack and Wall’s theory created the foundations for increasingly sophisticated models of pain perception, involving both bottom-up and central, context-sensitive pathways. In a chapter on the subject the neuroscientist Howard Fields describes current models of neurological pathways of pain perception and then considers its signalling functions—its meaning for the person, using the example of inadvertently touching a hot iron [22] (pp. 44–45):

Turning to the subjective experience: there are three distinct components [...] First, there is the purely discriminative part that includes recognizing the quality of the sensation as a burn and localizing it to your hand. Second, there is the motivational aspect associated with the desire to pull your hand away or to terminate the sensation. Third, there is an evaluative component *the thought of the damage that has been done to your hand and how that will affect your life in the hours and days ahead.* (italics added)

It is this third component—italicised in the quotation above—that is of special interest here, because we take it to involve: the thought of the impact of the damage to oneself, one’s agency and way of life, including its always important social aspects. In short, this evaluative component that is central to the experience of pain is thoroughly biopsychosocial. The ‘biopsychosocial’ appears here in the *intentionality* of pain, i.e. in what it is ‘about’, its meaning or representational content, which is, briefly: threat of loss of significant biopsychosocial function.

These considerations also provide a way of comparing ‘physical pain’ with ‘psychological pain’, or ‘distress’. As a simplification, psychological pain or distress is high on negative thoughts and feelings about one’s prospects. Fields has this [22] (p. 46):

In addition to its role in pain perception, the limbic system mediates emotional responses to a variety of factors including personal loss, anticipation of harm, and so on. The dysphoric states such as depression and anxiety share limbic system circuits with somatic pain. It is thus no accident that the word “pain” is often used to denote emotional pain that has no somatic component.

The evaluative component in pain perception figures prominently in psychological models of pain and distress, with clinical applications. Highly negative (fearful) appraisals about the effect of damage, or other negative events, and indeed of the pain or distress itself, on one’s future life are sometimes referred to as *catastrophizing*, and they typically risk having the effect of amplifying the experience. These models have the implication that psychological management of pain and distress should target among other things reduction of catastrophising [23].

A further psychosocial aspect of pain and distress is their function in interpersonal signalling and regulation of behaviour. This aspect is already implied in Wittgenstein’s account of pain expression: expressions of pain, behavioural, facial and verbal, induce caring responses from others. It has been theorised in various ways, for example in the Social Communication Model of Pain [24] and in evolutionary theoretic terms [25].

### 4.3 LOCATING CAUSES IN BIOPSYCHOSOCIAL SYSTEMS

#### *Identifying Dysfunctions and Modifiable Causes*

While disease is contextualised in the person as a whole, the immediate question is where the dysfunctional process is located: which system within the whole is dysfunctional, causing problems for the whole? The methodological assumption of healthcare is that the person as a whole is in trouble because of some dysfunctional part, a dysfunctional subsystem within the body/mind. It underlies the traditional individual focussed medical model of identification of clinical syndromes and diagnoses, and models of psychological processes in the individual that give rise to distress and activity limitations. The scientific details are in the medical and clinical psychological textbooks and will not be taken further here. We focus more on the broader implications of systems theory and the biopsychosocial approach. The main issue is that systems theory

envisages (causal) interactions everywhere, including within and between the organism and the environment, in which context the question arises: what is the logic of attributing causes of dysfunction to the organism rather than the environment?

In systems theory, one cannot begin to talk about the function of systems without reference to their operating environments. All biological systems function in interaction with others and ultimately in relation to fitness of the whole organism in a range of environmental conditions. One broad kind of pathway to dysfunction of the whole is poorness of fit between the expected environment, to which behaviour is adapted, and the actual environment, to which it isn't. This general point applies in the evolutionary context, for example in the so-called 'thrifty phenotype' hypothesis applied to obesity. Hales and Barker [26] hypothesise that in poor dietary conditions in utero metabolic mechanisms are set to maximise fat storage in expectation of subsequent, poor post-natal dietary intake; this mechanism would be highly adaptive in environments where poor dietary conditions in utero were reliably followed by poor dietary conditions postnatal, as may be reasonably assumed to be the case in our original evolutionary environments; but if this association breaks down, as in postnatal dietary environments that are actually high in accessible sugar/fats, the consequence would be a hard to modify tendency to excessive fat storage. The general idea of poorness of fit of previously adaptive mechanisms to later environments has an ontogenetic version in the learning theories in psychological science: behaviour shaped up by one set of environmental contingencies may be maladaptive in a subsequent environment. For example, if toddlers are reared in parenting styles involving multiple and conflicting commands, the child is likely to learn to ignore them and to seek to satisfy their own goals regardless, but this behaviour pattern will likely lead to poor outcomes in the classroom [27, 28].

Notwithstanding these considerations, we still locate the problem—the dysfunction—in the person. An obvious reason for doing so is the centrality of the person: it is the person to whom harm accrues—who suffers pain, distress, significant impairment of agency and loss. However, this consideration alone is superficial in a systemic context, because the cause may still lie outside the person, being done to, and this is the force behind the social model of disability which we have already had occasion to cite in this section. A more promising key to this issue is the one increasingly found in the health literature: the concept of

identifying modifiable causes. The idea is that, among all the criss-crossing causal pathways, what needs to be identified are promising targets for intervention. This approach is consistent with the interventionist approach to causality referred to in Chapter 3, Sect. 3.1, and is well suited to healthcare as applied science, seeking to change things, for the better. From this point of view, dysfunction attribution is in part—and somewhat paradoxically—shorthand for belief about promising possibilities for change. While ‘dysfunction’ and its cognates connote deficit, promising possibilities for change are opportunities.

In complex systems where there is a poor fit between the person’s behaviours and the environment, the question arises: where is potential for change? In conditions of the person that are lifelong, not amenable to change, the potential for change lies elsewhere, not in the person, but in social attitudes and resources—and this is a compelling argument of the social model of disability. For acquired long-term health conditions, it is likely that optimal outcomes from the person’s point of view come from a combination of—in no order—available high-quality healthcare, self-management, social support, plus non-discrimination by broader society.

Effective treatment or prevention technologies rely on targeting a cause of large enough effect, i.e. a causal factor identified in group studies that accounts for a large proportion of the outcome variance. The main point for the present purpose is that there are few causes of currently common health conditions with so large an effect that targeting them leads to complete prevention or complete cure, and for the majority, a multimodal approach to multiple factors is required.

These issues relate to the problem of *reduction of disease to a single primary cause*. If a normal function of a biological system is carried out by only that one system, then the failure of that function will be reducible to processes within that one system. For example, insofar as it is only the cardiovascular system that delivers oxygen to cells, failure to achieve that functional end, depleted oxygen delivery to cells, is attributable only to—and in this sense is reducible to—cardiovascular dysfunction, to a cardiovascular disease process such as atherosclerosis. In many cases however, and this may be the general rule, biological functions are affected by multiple subsystems, with the effect that achievement of a particular function is not a matter of processes in any one system, but may be affected by many interacting systems. However, such factors may not be relevant to the disease process once onset: for example, advanced

arteriosclerosis is not likely to be reversed by social policy affecting dietary changes, and quite different interventions may be needed, such as bypass surgery. On the other hand, for acquired, chronic health conditions, there is typically ongoing interaction with environmental, psychological and social factors. There is, in brief, no reduction to a primary cause, biological, psychological or social, but rather multiple systems of all kinds are involved at varying stages, some contributing risk for poor outcomes, others contributing to protection, including restorative and compensatory mechanisms.

### *Identifying Causal Mechanisms*

In setting the scene for developing biopsychosocial theory in the first chapter, we noted that evidence of biological, psychological and social causal factors in many health conditions comes from group statistical data in controlled study designs. The inference to causation relies on the empiricist approach to causation, after Hume and Mill, as association or correlation determined in experimental or quasi-experimental study designs. We noted however that such data in themselves provide no theoretical account of what kind of thing the variables stand for, or what kind of causal properties they have, separately or in combination. This absence of ontological-causal theory is particularly noticeable given the long-standing assumptions that physical(-chemical) processes alone are causal, covered by inviolable physical laws, that therefore biological factors can be causal only because biology is physics and chemistry, while mental events are scientifically odd epiphenomena, and social processes can hardly be conceptualised at all within this particular world view. The task of Chapter 3 and this chapter is to elucidate a general biopsychosocial theory capable of comprehending biopsychosocial data.

Issues raised by untheorised statistical data appear in the philosophy of medicine literature in discussion of whether the empiricist approach to determining causes, using controlled study designs and associated statistical methods, is sufficient, or whether it is also necessary to identify *causal mechanisms*. Federica Russo and Jon Williamson have proposed [29] (p. 158):

The health sciences infer causal relations from mixed evidence: on the one hand, mechanisms and theoretical knowledge, and, on the other, statistics

and probabilities. Statistics are used to show that the cause makes a difference to the effect, and mechanisms allow causal relationships to explain the occurrence of an effect.

This proposal turns on what a ‘causal mechanism’ is, and specifically on whether identifying a causal mechanism is distinct from determining probabilities in controlled study designs. In their paper cited above Russo and Williamson give examples of causal mechanisms, and in a related paper, Illari and Williamson [30] (p. 1) cite a definition of ‘causal mechanism’ from Machamer et al. [31]: ‘entities and activities organized in such a way that they are responsible for the phenomenon’—which is somewhat vague with more than a hint of circularity. A more informative definition is provided by Glennan [32] (S 344):

A mechanism for a behaviour is a complex system that produces that behaviour by the interaction of a number of parts, where the interactions between parts can be characterised by direct, invariant, change relating generalisations.

This is more informative, but is so exactly because it reintroduces the importance of invariant generalisations consistent with the empiricist approach to causation.

So what is a causal mechanism over and over what is established by controlled experimentation? The life sciences deal with complex systems changing over time, with probabilistic associations between inputs and outputs that are separated spatially by the inner workings of the system and by time. Confidence in having identified a causal mechanism is raised when the events are proximate, with fewer or no intervening processes and closer in time. So one idea behind ‘causal mechanism’ is just that we *fill in the intervening steps*, spatial and temporal, finding causal connections of ever closer proximity between inputs and effects, between, for example, environmental exposures at one time and poor health at a later time. This approach to ‘causal mechanism’—filling in the intermediate steps—is suggested by Illari and Williamson [30], and supported by considerations in, for example, Kincaid [33]. It is, however, readily accommodated in the empiricist approach to causation, a point well-argued by Kendler and Campbell [34]. A corollary of filling in the intervening steps, with ever closer proximity of links in the causal chain, taking into account other proximate factors at each step, in effect increases

probabilities from lower to higher. Further, as probabilities of association and correlations approach 1, the sample size required for confidence in generalisability reduces: a few well-designed, replicated experiments with relatively small samples will do. All this can be understood in terms of the empiricist approach.

There is however the famous limitation of empiricist epistemology, whether in the knowledge of causes or knowledge generally, namely, that it omits *theory*, envisaging knowledge by observation only. In the statement of their thesis quoted above Russo and Williamson bring together '(causal) mechanisms and theoretical knowledge' contrasted with statistics and probabilities, but it is worth distinguishing them. It is true that experimental method or approximations to it only ever establish correlations and associations, albeit generalisable and counter-factual that can support intervention to make a difference. The theory goes further, however, explaining why the correlations exist and why the intervention works, 'explaining' in the sense of fitting into a more or less well-established body of knowledge. On the other hand, it should be said that the additional need for theory is not a totally different requirement compared with establishing causal connections; rather, theory is typically a broader class of causal connections, themselves established or confirmed using experimental methodologies or approximations. So, if we elucidate the concept of 'causal mechanism' in terms of 'theory', just as if we elucidate it in terms of intervening steps, the process of identifying a causal mechanism and identifying a cause are similar—and specifically, identifying causal mechanisms is not a separate epistemological route to establishing causes. This conclusion is consistent with Alex Broadbent's discussion of these issues in epidemiology [35], and with Bert Leuridan's and Erik Weber's discussion of mechanistic evidence and the International Agency for Research on Cancer (IARC) [36].

As to what theorised biopsychosocial causal mechanisms are, the general concepts and principles have been the main topic in preceding chapters, including *systems, structures or forms, functioning towards ends, information and communication, coding, regulation and control*. And for causal mechanisms responsible for breakdown, their relevant negations, such as *error, abnormal, dysfunction and dysregulation*. These are the general concepts and principles; the specifics are diverse, depending on which system, vulnerable structures, ends and control mechanisms are being modelled.



### *Stress as a Biopsychosocial Causal Mechanism*

One of the most theorised and researched general causal mechanism for explaining biopsychosocial impacts on health is *stress*. Chronic stress, specifically, is hypothesised to be a key mechanism leading to stress-related biological reactivity such as inflammatory responses that adversely affect the immune system and other organs, raising the risk for a range of health conditions [37–40]. As an example of recent work in this area, here are summary statements from a recent major longitudinal study in *The Lancet* on the relation between resting amygdalar activity and cardiovascular events [41] (p. 2):

Chronic stress carries an attributable risk for cardiovascular disease that is on par with other recognised risk factors, such as smoking, increased lipid concentrations, hypertension, and diabetes. Despite the prevalence and potency of this risk factor, little is known about the mechanisms that translate stress into cardiovascular disease events... Our study provides several observations that together define a mechanism linking stress to cardiovascular events..., specifically that the amygdala could be a key structure in the mechanism... and that upregulation of haemopoietic tissue activity and increased atherosclerotic inflammation are additionally implicated in a neural–haemopoietic–arterial axis.

In the experimental psychology literature spanning animal and human research, the concept of stress is closely linked to fear, anxiety and depression. Stressors are various kinds of (perceived) threat, but with the specific feature of (perceived) *uncontrollability*. This idea has a long history in learning and personality psychology, for example, in Rotter’s locus of control theory [42] and Seligman’s learned helplessness model of depression [43]. Here is a formulation by Richard Lazarus [44] (p. 58):

A good way of thinking about stressful person–environment relationships is to examine the relative balance of forces between environmental demands and the person’s psychological resources for dealing with them. If the environmental load substantially exceeds the person’s resources, a stressful relationship exists... In psychological stress, the comparison is between the power of the environmental demands to harm, threaten, or challenge, and the psychological resources of the person to manage these demands... From the standpoint of this way of thinking, stress is particularly powerful

when the individual must struggle with demands that cannot easily be met... If the ratio of demands to resources becomes too great, we are no longer talking about high stress but trauma... The person feels helpless to deal with the demands to which he or she is exposed, and this can result in feelings of panic, hopelessness, and depression.

Psychologically, stress arises from exposure to salient negative, uncontrollable events, jeopardising the sense of agency. Salience covers what is essential to our biological and psychological life. Further, psychological stress essentially involves social factors such as task demands and access to resources. At the same time psychological stress is also biological, physiological: it is the activation of the arousal system, preparing for action to achieve important goals—but if the goals cannot be achieved, ever, or never enough, or never reliably, the arousal system is chronically active, and it is this chronic (hyper-) activity of the arousal system that is hypothesised to be the source of long-term biological damage. In short, chronic stress as the key hypothesised mechanism linking psychosocial factors with poor physical and mental health outcomes is—as to be expected—a mechanism that explicitly addresses criss-crossing biological, psychological and social processes. Key features of the hypothesised chronic stress mechanism are aspects of the core features we have proposed for biopsychosocial theory: the psychological sense of agency and action itself are compromised, raising risk for mental health problems, because social task demands are excessive and social resources inadequate, and the consequences of this chronic psychosocial misfortune is top-down dysregulation of critical biological processes raising risk of physical health problems.

### *Biopsychosocial Research Framework*

The biopsychosocial model, like the narrower biomedical model, is not a scientific theory or summary of scientific findings, but could be applied as a framework for organising and planning research. The N.I.M.H. Research Domains Criteria RDoC framework for mental health is a rigorously worked out example of a framework in this sense [45]. The RDoC framework is a 2-dimensional grid: the columns are for biological, psychological and social factors—in this sense, the RDoC framework is explicitly biopsychosocial—and the rows refer to specific neural-psychological-behavioural systems (such as *fear* and *reward* systems). The cells

can accommodate what is assumed known with more or less confidence, or could indicate what remains unknown, either not yet investigated or with mixed or inconclusive findings. We consider the RDoC research framework here because it is the best current and because it could be elaborated in various ways to have broader scope appropriate for the biopsychosocial model, for example application to physical health as well as mental health, incorporating biomedicine, and inclusion of a wider range and number of factors known to effect health and disease at various stages. Here are some main points that would be involved in such an elaboration:

- *Extension to physical health:* A research framework of this sort could be to apply to biological systems below as well as above the neck, to include such as the cardiovascular system as well as the central nervous system, in effect incorporating biomedicine, and potentially then able to have relevance to physical as well as to mental health conditions. Importantly, it would be able to accommodate the many kinds of pathways and conditions that do not fit neatly into either of these two categories, such as risks involving chronic stress, or the so-called psychosomatic conditions or medically unexplained symptoms. The expanded framework would in effect have the advantage of recognising interactions between the brain and other biological systems, and hence be able to accommodate the emerging evidence outlined in Chapter 1, Sect. 1.2, implicating psychosocial factors in the aetiology and course of medical conditions. Assuming the grid has explicit relevance to both systemic function and dysfunction in the rows, on which more below, some aspects of research findings relevant on mental health on the one hand and physical health on the other would diverge significantly, for example, confirmation of primary biological progressive disease mechanisms and treatments in some physical diseases. But in other areas of the grid, particularly relating to aetiological risk factors accumulating through the lifespan, or in areas of the grid—to be proposed for addition below—on management of chronic conditions and factors affecting quality-of-life, similarities among mental and physical health conditions would be more apparent.
- *More discrimination among kinds of psychological and social factors relevant to health and disease.* For example, to accommodate aspects of agency: agency as perceived, and agency related to social

factors including task demands (e.g. work; dependents) and access to resources and opportunities (these of many kinds, including access to treatment). This would require more columns. The RDoC framework is work in progress, adaptable as the science develops; current versions have around 5 columns for biological factors, around 1 for psychological, and around 0–1 for social factors (e.g. [45, 46]).

- *Acknowledgment of non-social environmental health risks*, especially important if physical health is included, again requiring more columns, to include factors such as ambient air quality and available diet.
- *Explicit specification of health conditions or ‘diseases’*, not only the biological systems. The RDoC framework at present has no explicit conceptualisation or characterisation of mental health conditions, connected with the aim of replacing current psychosocial diagnostic criteria with biological criteria [47–49]. To accommodate specification of health conditions, probably a third dimension of the grid would be needed. This is easier to see if we imagine the grid incorporating biomedicine in which the issues are better worked out: the rows would be specific systems such as the immune system, with implicit reference to its components and functions, and the columns would specify factors affecting functioning, but probably a third dimension of the grid, distinct though theoretically closely connected to the rows and the columns, would be needed to specify the dysfunctions and disorders of the immune system. The cells in this now 3-dimensional grid could then accommodate findings of the specific subsystems responsible for harmful health conditions warranting healthcare attention.
- *Need to discriminate among stages of health conditions* would arise once a dimension specifically for health conditions was explicitly in place. It would also be necessary, as we have emphasised previously, to distinguish between research questions referring to, first, aetiology of disease incidence, in population samples; second, disease progression or maintenance, in patient samples; and third, factors affecting quality-of-life in chronic conditions. These discriminations are necessary because they are distinct research questions, requiring distinctive methodologies and sampling, and also—especially relevant to our main theme—because the balance of biological, psychological and social involvement can vary substantially depending

on the stage of a condition. For example, for cardiovascular disease biological processes dominate as maintaining factors and targets for intervention such as surgery in the advanced stages of the disease; whereas, if the question is the aetiology of cardiovascular disease, accumulation of risk factors in the population, to be applied as basis for prevention technologies, or application to advising an at-risk individual patient, then lifestyle social and factors, such as exercise and time of access to treatment, figure large. And for chronic diseases, in fact for all diseases where the person is alive and managing, not in coma, there are always issues of agency and the quality of life. A framework for organising or planning research into management of chronic diseases would, therefore, need to accommodate the full range of biological, psychological, environmental and social factors.

- *Population level as opposed to individual level questions*, for example, incidence vs. susceptibility, might require different grids, the one to do with differences between individuals (types), the other differences between populations. The UK NICE conceptual framework for public health [50], for example, distinguishes between individual and population patterns of disease and their causal mechanisms; both include biological, social and related factors, but the latter has additional interactions with a range of other factors including political and economic.
- Finally, a further dimension of variation is *developmental*. All biological and psychological systems in health and disease have developmental trajectories, within which there is variation in the relative influence of biopsychosocial and environmental factors, including in factors affecting vulnerability and resilience to adversities and illness. Hence all the research questions would have to allow for age variation.

The points above indicate what, based on considerations so far in this essay, would constitute an adequate framework for organising health research and identifying areas of relatively certainty and important unknowns. Possibly further dimensions could be added, for example on ‘impact’, estimating the relative importance of knowing more about a specific health condition at a specific stage, for treatment or prevention, depending on, for example, prevalence, projected prevalence, among what age-group, healthcare costs, associated cross-sectoral costs, etc. But, in any case, the elaborated framework as sketched above is already

multidimensional, needs far more than a two-dimensional grid, can hardly be represented diagrammatically, though could be split up into different diagrams, but it is bound to be complicated if able to accommodate and organise the entire basic and clinical science health research—this can hardly be expected to be simple. In practice of course such a multidimensional monster grid to organise biopsychosocial research across the whole of healthcare will never be constructed because too big, too complicated and of no practical use. Small segments of the hypothetical framework are written up in reviews for circumscribed specifics: for some conditions, some treatments, some stages, some health economic analyses, some policies, other angles. Otherwise, it exists in the scientific literature as a whole, broadly construed, across the range of biopsychosocial and environmental sciences applied to health.

In the next section, we consider the tension—intrinsic to healthcare as it has developed over the past few decades—between research data of the kinds considered above, on groups, and clinical care of the individual.

### *Clinical Epistemology*

It is something of an irony that while health research has made such strides over the past few decades, while knowledge has increased, certainty in the clinic is just as likely to have gone down as up! This is connected with the fact that much of what has been discovered is about complex, multifactorial causation. We know more about the body and mind, their functions and dysfunctions, and their interaction with the environment, and more about the treatment of biological and psychological health problems, but this has come along with increasing appreciation of the complexity of the problems, beyond physically damaged tissue or biological infection, involving multiple interacting biological systems, along with increased understanding of regulating systems with wide interactive reach, including the central nervous system and psychological functioning. Linked with multifactorial complexity, this new complicated knowledge is statistical, based on group studies, delivering associations such as relative risk and odds ratios, and quantifications of treatment effects such as effect size and number-needed-to-treat—and how all these statistics relate to a particular patient is unsettled. Notwithstanding the benefits of evidence-based practice, the challenges of inference from population-based aetiological risk studies and clinical treatment trials to preventative management and treatment of the

individual patient are significant [51, 52]. The challenges here do not disqualify the application of the experimental method, following Mill's methods of agreement and difference, or approximations to them, to determine causes and effects (as outlined in Chapter 1, Sect. 1.3, under the heading “[Biopsychosocial Data in Search of Theory](#)”). There are in fact no other serious players on this particular pitch. Application includes reliance on randomised controlled treatment trials (or better, meta-analyses of multiple such trials) as being the most logically valid way of identifying treatment effects (see e.g. [53]). Experimental method can result in reliable positive findings, but also, and of high importance, reliable negative findings, likelihoods of no or no clinically significant effect of a treatment, compared with no treatment, or with a harmless placebo. The epistemological problem is not how to establish that a treatment technology has some effect or no effect in group samples, but rather that, because of many kinds of complexity (in the condition, in the sampling, in individual differences), treatments are rarely effective for all individuals, and application of the data to care of the individual patient is not straightforward. For this, as is often said, thorough assessment and clinical judgement are needed to combine with knowledge of the basic and clinical science.

Complexity and uncertainty are most marked where there is evidence of causation by multiple factors of small effect. Conversely, simplicity and certainty are most marked where single, primary causative factors are presumed. The single, primary factor approach works well in some specialist areas of biomedicine and psychological therapy, and less well in clinical settings with caseload is not restricted to a narrow range of conditions, in settings such as primary care, palliative care, care of the elderly, and community mental health. Other limitations of the single factor approach are apparent in medical wards and outpatient clinics in which some patients present with pain, distress and activity limitations in the absence of biomedically determined conditions. All these contexts require a broader causal theory, more complex, about which much is unknown at present, to do with biological/psychological/social interactions of the sort being explored for example in the chronic pain, health psychology and public health literatures.

This new complexity creates much uncertainty, in clinicians, patients and students. It can be resolved by a dogmatic certainty that the real cause must be one or another sort—something biological, psychological or social—though at the cost of selective inattention to other factors,

the risk of over-reliance on one treatment approach, and detachment from anomalies. More adaptively, the uncertainty has to be tolerated. The more responsibility a clinician has, the more obligation they have to know the science as well as the patient and to keep a mind open to complexity and alternatives, at the same time as needing to make definite decisions and recommendations one way or the other.

There are occasions, in response to questions from patients or from students as to causes, in complex cases, where single aetiology of large effect has been excluded, and the picture looks more like multiple aetiology of small effect, it is as correct as anything else to say: 'it is a complex biopsychosocial picture'. This move is by all means somewhat vague and hand-waving, connected to the criticism of the biopsychosocial model reviewed in the first chapter, that it is vague and too often used for unhelpful hand-waving. Engel's model has stood ready to accommodate emerging findings of biopsychosocial complexity, and being so accommodating has made it hard to capture in a few words except vaguely. However, complexity and uncertainty have come from the science; they are not peculiar features of a model—no point blaming the messenger. And, in fairness to the biopsychosocial model, the generalised single primary cause models are the same—vague hand-waving to everything being biological/biomedical, or else all psychological, or social. The science of the past few decades has all but ruled out these single primary cause general models, and endorsed the broader biopsychosocial approach. The broader approach is also able to be more discriminating, more empirically based than the previous generalised single factor models. Biological, psychological and social factors may be involved in specific health conditions, at specific stages, but whether they are or not, and in what degree, is not known in advance, but only by doing the science.

#### 4.4 COMPARE AND CONTRAST PHYSICAL AND MENTAL HEALTH CONDITIONS

##### *Psychiatry and 'The Rest of Medicine'*

Psychiatry is obviously psychological, at first glance, but also obviously social, at second glance, while biological to a degree, while the rest of medicine—according to the biomedical model—manages well enough with the biological only. In this sense, the question of psychiatry's relation to 'the rest of medicine' stands proxy for the rationale and validity



of the biopsychosocial model—and vice versa. In fact, Engel chose just this issue as the starting place for his 1977 paper; he turned on its head the aspiration for psychiatry to emulate the rest of medicine, recommending the opposite: make the rest of medicine more like psychiatry—more psychosocial, not biological only [4] (p. 129).

On the other hand, as we noted at the beginning of the first chapter (under the heading “[The Presumed ‘Overarching Framework’](#)”), the ‘rest of medicine’ is not one thing, and the various medical specialities differ in their relative involvement with biological, psychological and social factors. Primary care (also known as general medical practice, or family medicine) is much involved with the psychosocial, as is public health, and palliative care, as well as many aspects of care on acute medical wards. In this sense, the contrast is not so much between psychiatry with the rest of medicine, but between psychiatry along with many other areas of medicine, contrasted with biomedicine. Taking these considerations things into account much qualifies the idea that psychiatry is so different because of its involvement with the psychosocial. Nevertheless, psychiatry can still be regarded as the odd one out compared with ‘the rest of medicine’, for reasons that go much deeper than detailed and discriminating considerations of the above sort about varying degrees of involvement with the psychosocial.

### *The Difference Is Deeply Theorised and Institutionalised*

The perceived difference between mental and physical health conditions and healthcare is underpinned by the great historical dichotomies outlined in the first chapter, Sect. 1.3 (under the heading “[Prejudicial Theory: Physicalism, Reductionism, Dualism](#)”): mind/body dualism, and the separation of the social and moral sciences from the natural sciences. Thomas Szasz’s highly influential 1960s critique of psychiatry [54] relied on these dichotomies. But worse, the two sides of the dichotomies were not equally balanced in respect of scientific validity, especially in connection with determining causes and interventions—matters fundamental to medicine. Rather, against the background of physicalist reductionism, which underpinned the dichotomies, as reviewed in Sect. 2.2, while physical health conditions involved recognised causes and effects, researchable and manageable by proper biological/biomedical science, mental disorders were something else altogether, barely recognisable let alone theorised, and psychiatry along with them.

Built on top of the historical dichotomies in deep theory are the reinforcing, maintaining effects of having organised the whole of healthcare training and delivery around physical health problems on one side of the road and mental health problems on the other. On one side, biomedicine performs best with biological mechanisms in physical diseases, and psychosocial involvement, if any, is out of scope. On the other side, theoretical or practical preoccupation with ‘mental abnormalities’ such as delusions and other hard to understand mental states and behaviour tends to neglect somatic signs and symptoms, and does not bring into focus people as a whole and their social circumstances. The dichotomy between mental and physical health conditions is historically theorised and currently institutionalised and practised.

### *The Biopsychosocial Model Highlights Similarities*

There are several reasons why the picture is changing however. Mental health conditions are more evident, linked to increasing public awareness and efforts to decrease stigma, and the extent of associated activity impairments such as days lost to work is better understood and increasingly recognised as comparable with those in physical health conditions. It is increasingly recognised that physical and mental health problems often co-occur, complicating each other, and therefore also complicating our healthcare system, given that it is currently organised on the basis of separating them out, along with the clinical expertise for managing them. And as regards aetiology, public health and prevention, recent epidemiology suggests that the two kinds of health problem can share aetiological risk factors, possibly implicating shared mechanisms. These social and scientific developments change policy, as for example in the UK NHS policy paper ‘No health without mental health’ [55]. In this section, we review these issues in more detail, with reference to the biopsychosocial theory and science set out in previous chapters.

Considering aetiology, we noted in the first chapter, Sect. 1.2, the emerging epidemiological evidence that implicates psychosocial as well as biological risk factors including genetic for many physical health conditions. It also suggests that some risks of all sorts are shared between some physical health conditions and some mental health conditions; it is not the case that risk factors divide neatly into those to physical health on the one hand and those to mental health on the other. Drilling into hypothesised mechanisms, we saw in section [Stress as a Biopsychosocial](#)

**Causal Mechanism** that chronic stress and its biological effects are commonly implicated in the aetiology of many physical and mental health conditions. Again, it is not the case that pathogenic mechanisms neatly divide between those for physical health conditions and those for mental health conditions. As corollary, preventative strategies and technologies, for many physical and mental health conditions, overlap. Public health does not have two unconnected tasks, one for physical health promotion and another for mental health promotion.

Post onset, especially for the long-term conditions, also considered in Sect. 1.2 under the heading “**Emerging Evidence of Psychosocial Causation**”, psychosocial factors affect biomedical management, in matters such as access and collaboration over management plan, for example ongoing medication; as well as affecting psychological adjustment and quality of social life. These diverse psychosocial issues coincide or at least overlap for both physical health and mental health long term conditions. We went on to note the connected finding that physical health problems raise risk for mental health problems and vice versa. The causal pathways are diverse, but include such as chronic physical ill-health imposes activity restrictions and loss of amenity, and pain, all of which raise risk of high anxiety and low mood; mental health chronic conditions can be associated with risk factors for physical health problems, such as social exclusion, poor diet, smoking, and higher thresholds for medical attention to physical health problems. The picture that emerges, therefore, is not that of patients with physical health problems, and an entirely different set of patients with mental health problems. All these considerations—regarding aetiology, adjustment, quality of life, and bidirectional complications—serve to break down the dichotomy between mental health conditions and physical health conditions. They highlight the importance of psychological and social as well as biological factors in health and disease, and they need broad biopsychosocial theory to accommodate them.

The general drift of the biopsychosocial systemic approach—as can be expected from its name—is to view physical and mental health conditions under a unified ‘health problem’ heading. The core common feature is a substantial negative effect on the person’s agency, associated with distress: with worry and fear about their safety and their future and their dependents.

In the broader biopsychosocial picture, the key secondary difference between physical health problems and health problems is that some but

not all physical health problems have a biomedically identifiable maintaining cause—a disease process or lesion—while this is probably not the case for mental health problems. This is a critical difference and it stands out most clearly for physical health problems that are biomedically well understood and treatable, in a relatively short timeframe, without therefore impacting on what is presupposed as an otherwise normal life. Cure of infectious disease by antibiotics, surgical interventions that are now routine such as hip replacements and even cardiac surgery, especially where all the psychological and social conditions for access, detection and intervention are in place, and which therefore can be ignored, stand out as triumphs of biomedicine. If we start with the underlying presumption that physical health problems are purely physical—and entirely different from mental health problems—these are the cases we will attend to, and we would tend to neglect the kinds and aspects of physical health problems that don't fit the picture: regarding aetiology, chronic conditions and comorbidities as reviewed briefly above. And, coming from the other direction, the assumption that mental health problems are quite different from physical health problems because exclusively to do with the mind, or the person, is also problematic. For example, some mental health conditions have some response to pharmacotherapy. It is true that psychological therapy is often indicated along with medication for mental health conditions, but equally, as is now being recognised, it is often indicated alongside medical management of physical health conditions [56]. As to mental health conditions, as opposed to physical health conditions, being integral to the personality, the contrast is less marked for long-term conditions of either type, as previously remarked in Sect. 4.2. Also, some mental health conditions such as obsessive-compulsive disorder are typically seen by the person as externally imposed, rather than as integral to themselves. This is probably the rule for mental health conditions rather than the exception. This is a complicated clinical area but the point, in short, is that only for a particular sub-class of mental health conditions is there a strong presumed link with personality, that is, the so-called 'personality disorders'.

Another way of viewing the similarities and differences between mental health conditions and physical health conditions is through the lens of the hypothetical virtual biopsychosocial research framework sketched above (Sect. 4.3). In addition to the specification of biological and neurological systemic functioning, this framework was imagined to include specification of health problems, physical and mental, and to have

complete coverage of stages, from risks of onset through to post-onset maintaining causal mechanisms, interventions, and factors affecting adjustment and quality-of-life in long-term conditions. The columns of the grid would include biological, psychological and social factors, and the cells research findings. The upshot of this is that the relative importance of biological compared with psychosocial factors would be most marked between mental and some physical health problems at just one—albeit very important—point, namely post-onset maintaining causal mechanisms and interventions. For some physical health problems, these would be mainly biological with little psychosocial. But for all other stages: aetiological pathways to onset, and post-onset adjustment and quality-of-life, the pattern of relative weights of biological, psychological and social would be evened out and would certainly not be all biological for all physical health problems, and all psychological and social for mental health problems.

#### 4.5 LOCATING THE BIOPSYCHOSOCIAL MODEL

We noted in the first chapter that the biopsychosocial model has been charged with *vagueness* in the clinic, as well as vagueness as a scientific theory and as a ‘model’. It is true that Engel wrote his 1980 paper [5] on clinical application of the model, giving rise to the reasonable inference that the biopsychosocial model was a guide to clinical practice. However, as we went on to note in Sect. 1.2, there have been many developments in the intervening decades with more direct relevance to scientific content and guiding clinical practice. There have been new research programmes to investigate the causes of diseases and disease mechanisms, and technologies for prevention, early detection and treatment. These, in turn, have led to treatment guidelines for specific conditions at specific stages, to the whole apparatus of evidence-based clinical care, to be used alongside a thorough assessment of the individual case. The point is simply that, given all this basic and clinical science of the last few decades, the biopsychosocial model cannot usefully be regarded as some additional statement of the science or as a tool to guide clinical decision-making. It is true the model advises us to keep one’s mind open to the range of biopsychosocial factors, but the treatment guidelines and the science behind them already now say this, if applicable, and we don’t need a general model to repeat the fact—especially not to repeat it vaguely instead of paying close attention to the science of specific conditions and stages.

While this may be a solution for the biopsychosocial model of the vagueness problem, it works only, as we noted in section “[So What’s the Point of a ‘General Model?’](#)”, by raising the more fundamental question: what is the point of having a general model at all? We then located the task of the general model as defining biopsychosocial ontology and causation, in Sect. 1.3, noting the special need for this because of the deeply entrenched assumptions of physicalism, dualism and reductionism that have been so influential in the development of the life and human sciences. With these assumptions, only physical properties and causation appear real, while the mind is a non-causal epiphenomena, and social organisation and processes can hardly be comprehended at all. In short, the scientific and philosophical back story is more or less entirely antithetical to theorising biopsychosocial ontology and interactions. Hence the need for a new general theory for this purpose. We pursued this, tracking the science, in Chapters 2 and 3.

In Chapter 2 on biology we used the approach especially suited for the present purpose that relates biological processes to physics. Life forms do extraordinary things with energy, holding up the general direction of the second law of thermodynamics, for a while, the key being control by genetic code, essentially prone to error, to doing it differently, making space for evolutionary diversification. The key ontological shift compared with physicalism is away from few primary physical qualities and laws, variations on the theme of energy and energy conservation, towards multiplicity and diversity of dynamical forms with their own distinctive principles of change and causal interaction, all however retaining consistency with the physics of the matter. The corresponding key epistemological shifts are from generality to specificity, simplicity to complexity, and from knowledge of inviolable facts to active knowing, something more like ‘trial and error’. Moving on from biology, the psychological and the social were considered in Chapter 3. The primary concept of the psychological is identified as agency. This connotes altogether: causation, in the sense of regulatory control, authorship, individual differences, and self-determination. Agency is thoroughly biological: it is embodied, and accordingly has to secure the conditions necessary for biological life, specifically those related to maintaining appropriate energy differentials. At the same time, agency for us social beings needs recognition in the social group, and generally assumes a socio-political dimension, connoted by the related concept of autonomy. The primary function of the social is

identified in the model as the regulatory control of the distribution of resources necessary for biological life, but also of resources and opportunities for psychological development and cultivation of agency.

In this biopsychosocial theory, concepts of health and disease appear in prototypical form at the beginning, in the differences between survival and non-survival of biological organisms, between a biological system's working or breaking down. The basic facts of biological health and disease carry through into the biopsychosocial whole, being joined by concepts of psychological health and ill-health, related to agency, and concepts of psychosocial health and disadvantage, marked by exclusion from social relationships, resources and opportunities. Causal pathways run within and between all these systems and the many subsystems that serve them, in health and ill-health. The exact pathways and the size of effects vary with the health condition, its stage, and the challenges it presents to the person as agent.

As well as major developments in the basic and clinical sciences since Engel's original papers, there have been other major developments in dedicated models of health and disease and clinical practice. Three such have been mentioned so far in this chapter, Sect. 4.2: the social model of disability, which contests attribution of cause of activity limitations to the person rather than to the ill-resourced, socially excluding environment; the model of patient-centred care, which locates the person as patient, their aims and values, at the centre of healthcare, and the Recovery model, which theorises the need of the person with a chronic health condition to recover their life notwithstanding. These dedicated models emphasise specific important aspects of healthcare that broadly relate to individual differences, the person, the broader social and political context, and managing with chronic conditions—typically with explicit contrast with a perceived simple and over-simple 'medical model', with its focus on biological disease processes in the individual. In this sense, these models have taken up challenges and tasks of the sort that Engel identified, but with more elaboration, depth and detail than the biopsychosocial model itself.

As proposed here, biopsychosocial theory and the biopsychosocial model define the conceptual foundations of a new approach to health, disease and healthcare, one that responds to the accumulating evidence implicating many and diverse processes of kinds indicated by the name, and more besides, particularly the physics and chemistry of our bodies

and the environment, at one end, and social and economic policy at the other. It is more general than the science of specifics, or single disciplines, or dedicated models of clinical care. It is more like a view of human nature, based in the current science, one that includes propensity to health and disease. As a view of human nature and its vulnerabilities, the biopsychosocial model is comparable to the biomedical model. The biomedical model has two versions however: the old version, running to approximately mid-twentieth century, assuming, as Engel saw, physicalist reductionism and dualism, the other brand new and going from strength to strength since, at the cutting edge of reconstructing the relationship between biology, physics and chemistry, and articulating new models involving not only the inviolable physics and chemistry of energy, but also vulnerable forms regulated to ends. The new research programmes have advanced biomedicine, but at a conceptual level they open up worlds beyond the biological to include the psychological and the social. This conceptual opening up is of huge importance given that the conceptual foundations of health science and healthcare need to be able to comprehend and respond to all the new findings on psychosocial factors that have been accumulating over the past few decades, on the social determinants of health, the effectiveness of psychological and social treatments, and the increasing prevalence of long-term health conditions.

Biopsychosocial theory, incorporating the psychosocial and the political, also involves morality. The biopsychosocial model of health and disease has conceptual connections with bioethics. This is a contrast with the biomedical model, in either its old or new forms. To the extent that the biomedical model embraced physicalist reductionism, it was not entitled to any normative concepts, not even the difference between health and disease, and definitely not morals. Normativity has no place in physics and chemistry. The new biomedical model that invokes regulatory control mechanisms has normativity, but so far restricted to internal somatic systems and does not yet comprehend the whole human being as an agent in the interpersonal, socio-political world. To have this reach, the biopsychosocial model is required, and the term 'bioethics' could be expanded to 'biopsychosocial ethics'. At the foundational level, all normativity is interconnected. The 4 principles of bioethics laid out by Beauchamp and Childress [57] employ terms and relations that are foundational in biopsychosocial theory: autonomy of the person, harm and benefits to the person, social distribution of resources. The biopsychosocial theory does not resolve ethical disputes but indicates their terms



and the friction points where they arise. Biological health, psychological health, autonomous exercise of agency and values, social provision of resources necessary for these things—are all goods from our point of view as biopsychosocial beings, but they can be hard to achieve together since they can come into conflict one with one another. The individual may come into conflict with family, clinicians or the law, over what is good for them; what is in the interests of the individual may conflict with what is in the interests of the community; attribution of ‘illness’ or ‘disability’ may have benefits in terms of access to healthcare and support, but it downgrades recognition of autonomy, with potential for harm; provision of resources can conflict with promoting agency; equal distribution of resources competes with individual and group interests. And there are boundary issues, for example as to when biological life becomes psychological life with moral value protected by law, or as to when psychological life has come to an end in severe brain damage while the biology continues. Biopsychosocial theory cannot resolve these many kinds of moral dilemmas, but their terms and the potential for conflicts over priorities and boundaries appear at its foundations.

#### REFERENCES

1. Jaspers, K. (1962). *Allgemeine psychopathologie* [General psychopathology] (J. Hoenig, M. W. Hamilton, trans.). Manchester: Manchester University Press. (Original work published 1913).
2. Shakespeare, T. (2006). The social model of disability. In L. J. Davies (Ed.), *The disability studies reader* (2nd ed., pp. 197–204). New York & Oxford: Routledge.
3. Oliver, M. (2013). The social model of disability: Thirty years on. *Disability & Society*, 28(7), 1024–1026. <https://doi.org/10.1080/09687599.2013.818773>.
4. Engel, G. L. (1977). The need for a new medical model: A challenge for biomedicine. *Science*, 196(4286), 129–136.
5. Engel, G. L. (1980). The clinical application of the biopsychosocial model. *American Journal of Psychiatry*, 137(5), 535–544.
6. Brody, H. (1999). The biopsychosocial model, patient-centered care, and culturally sensitive practice. *The Journal of Family Practice*, 48(8), 585–587.
7. Creed, F. (2005). Are the patient-centred and biopsychosocial approaches compatible. In P. D. White (Ed.), *Biopsychosocial medicine: An integrated approach to understanding illness* (pp. 187–199). New York: Oxford University Press.

8. Smith, R. C., Fortin, A. H., Dwamena, F., & Frankel, R. M. (2013). An evidence-based patient-centered method makes the biopsychosocial model scientific. *Patient Education and Counseling*, 91(3), 265–270. <https://doi.org/10.1016/j.pec.2012.12.010>.
9. Evers, A. W. M., Gieler, U., Hasenbring, M. I., & van Middendorp, H. (2014). Incorporating biopsychosocial characteristics into personalized healthcare: A clinical approach. *Psychotherapy and Psychosomatics*, 83(3), 148–157. <https://doi.org/10.1159/000358309>.
10. Davidson, L., & Strauss, J. S. (1995). Beyond the biopsychosocial model: Integrating disorder, health, and recovery. *Psychiatry*, 58(1), 44–55.
11. Frese, F. J., III, Stanley, J., Kress, K., & Vogel-Scibilia, S. (2001). Integrating evidence-based practices and the recovery model. *Psychiatric Services*, 52(11), 1462–1468.
12. Slade, M. (2009). *Personal recovery and mental illness: A guide for mental health professionals*. Cambridge: Cambridge University Press.
13. Bennett, B., Breeze, J., & Neilson, T. (2014). Applying the recovery model to physical rehabilitation. *Nursing Standard*, 28(23), 37–43.
14. Boorse, C. (1975). On the distinction between disease and illness. *Philosophy & Public Affairs*, 5(1), 49–68.
15. Boyd, K. M. (2000). Disease, illness, sickness, health, healing and wholeness: Exploring some elusive concepts. *Med Humanit*, 26(1), 9–17.
16. Osler, W. (2009). *Aequanimitas and other addresses*. San Francisco, CA: Internet Archive (Original work published 1904).
17. Huber, M., Knottnerus, J. A., Green, L., Horst, H., Jadad, A. R., Kromhout, D., et al. (2011). How should we define health? *BMJ*, 343. [<https://doi.org/10.1136/bmj.d4163>].
18. Wittgenstein, L. (1953). *Philosophical Investigations*. (G. E. M. Anscombe, Trans.). Oxford: Blackwell.
19. Hardcastle, V. G. (1999). *The myth of pain*. Cambridge, MA: The MIT Press.
20. Bourke, J. (2014). *The story of pain: From prayer to painkillers*. Oxford: Oxford University Press.
21. Melzack, R., & Wall, P. (1965). Pain mechanisms: A new theory. *Science*, 150(3699), 971–979. <https://doi.org/10.1126/science.150.3699.971>.
22. Fields, H. L. (2007). Setting the stage for pain: Allegorical tales from neuroscience. In S. Coakley & K. K. Shelemay (Eds.), *Pain and its transformations: The interface of biology and culture* (pp. 36–61). Cambridge, MA: Harvard University Press.
23. Quartana, P. J., Campbell, C. M., & Edwards, R. R. (2009). Pain catastrophizing: A critical review. *Expert Review of Neurotherapeutics*, 9(5), 745–758. <https://doi.org/10.1586/ern.09.34>.
24. Craig, K. D. (2009). The social communication model of pain. *Canadian Psychology*, 50(1), 22–32.

25. Williams, A. C. C. (2002). Facial expression of pain: An evolutionary account. *Behavioral and Brain Sciences*, 25(4), 439–455.
26. Hales, C. N., & Barker, D. J. P. (2001). The thrifty phenotype hypothesis: Type 2 diabetes. *British Medical Bulletin*, 60(1), 5–20.
27. Patterson, G. R., DeBaryshe, B. D., & Ramsey, E. (1989). A developmental perspective on antisocial behavior. *American Psychologist*, 44(2), 329–335.
28. Patterson, G. R. (1982). *Coercive family process*. Eugene, OR: Castalia.
29. Russo, F., & Williamson, J. (2007). Interpreting causality in the health sciences. *International Studies in the Philosophy of Science*, 21(2), 157–170. <https://doi.org/10.1080/02698590701498084>.
30. Illari, P. M., & Williamson, J. (2011). Mechanisms are real and local. In P. M. Illari, F. Russo, & J. Williamson (Eds.), *Causality in the sciences* (pp. 818–844). Oxford: Oxford University Press.
31. Machamer, P., Darden, I., & Craver, C. F. (2000). Thinking about Mechanisms. *Philosophy of Science*, 67(1), 1–25.
32. Glennan, S. (2002). Rethinking mechanistic explanation. *Philosophy of Science*, 69(S3), S342–S353.
33. Kincaid, H. (2011). Causal modelling, mechanism, and probability in epidemiology. In P. M. Illari, F. Russo, & J. Williamson (Eds.), *Causality in the sciences* (pp. 70–90). Oxford: Oxford University Press.
34. Kendler, K. S., & Campbell, J. (2009). Interventionist causal models in psychiatry: Repositioning the mind-body problem. *Psychological Medicine*, 39(6), 881–887. <https://doi.org/10.1017/S0033291708004467>.
35. Broadbent, A. (2011). Inferring causation in epidemiology: Mechanisms, black boxes and contrasts. In P. M. Illari, F. Russo, & J. Williamson (Eds.), *Causality in the sciences* (pp. 45–69). Oxford: Oxford University Press.
36. Leuridan, B., & Weber, E. (2011). The IARC and mechanistic evidence. In P. M. Illari, F. Russo, & J. Williamson (Eds.), *Causality in the sciences* (pp. 91–109). Oxford: Oxford University Press.
37. Steptoe, A. (2005). Remediable or preventable psychological factors in the aetiology and prognosis of medical disorders. In P. White (Ed.), *Biopsychosocial medicine* (pp. 59–75). Oxford: Oxford University Press.
38. Lightman, S. (2005). Can neurobiology explain the relationship between stress and disease. In P. White (Ed.), *Biopsychosocial medicine* (pp. 103–111). Oxford: Oxford University Press.
39. Buric, I., Farias, M., Jong, J., Mee, C., & Brazil, I. A. (2017). What is the molecular signature of mind-body interventions? A systematic review of gene expression changes induced by meditation and related practices. *Frontiers in Immunology*, 8(670). <https://doi.org/10.3389/fimmu.2017.00670>.
40. Slavich, G. M., & Irwin, M. R. (2014). From stress to inflammation and major depressive disorder: A social signal transduction theory of depression. *Psychological Bulletin*, 140(3), 774–815.

41. Tawakol, A., Ishai, A., Takx, R. A. P., Figueroa, A. L., Ali, A., Kaiser, Y., et al. (2017). Relation between resting amygdalar activity and cardiovascular events: A longitudinal and cohort study. *The Lancet*, 389(10071), 834–845.
42. Rotter, J. B. (1966). Generalized expectancies for internal versus external control of reinforcement. *Psychological Monographs: General and Applied*, 80(1), 1–28.
43. Seligman, M. E. P. (1972). Learned helplessness. *Annual Review of Medicine*, 23(1), 407–412.
44. Lazarus, R. S. (1999). *Stress and emotion: A new synthesis*. London: Free Association Books.
45. Insel, T., Cuthbert, B., Garvey, M., Heinssen, R., Pine, D. S., Quinn, K., et al. (2010). Research domain criteria (RDoC): Toward a new classification framework for research on mental disorders. *American Journal of Psychiatry*, 167(7), 748–751.
46. Cuthbert, B. N. (2014). The RDoC framework: Facilitating transition from ICD/DSM to dimensional approaches that integrate neuroscience and psychopathology. *World Psychiatry*, 13(1), 28–35.
47. Wakefield, J. C. (2014). Wittgenstein’s nightmare: Why the RDoC grid needs a conceptual dimension. *World Psychiatry*, 13(1), 38–40.
48. Phillips, M. R. (2014). Will RDoC hasten the decline of America’s global leadership role in mental health? *World Psychiatry*, 13(1), 40–41.
49. Bolton, D. (2017). Clinical significance, disability and biomarkers: Shifts in thinking between DSM-4 and DSM-5. In K. Kendler, & J. Parnas (Eds.), *Philosophical Issues in Psychiatry IV. Psychiatric Nosology* (pp. 8–16). Oxford: Oxford University Press.
50. Kelly, M. P., Stewart, E., Morgan, A., Killoran, A., Fischer, A., Threlfall, A., et al. (2009). A conceptual framework for public health: NICE’s emerging approach. *Public Health*, 123(1), e14–e20. <https://doi.org/10.1016/j.puhe.2008.10.031>.
51. Evidence-Based Medicine Working Group. (1992). Evidence-based medicine. A new approach to teaching the practice of medicine. *Journal of the American Medical Association*, 268(17), 2420.
52. Greenhalgh, T., Howick, J., & Maskrey, N. (2014). Evidence based medicine: A movement in crisis? *British Medical Journal*, 348, g3725.
53. Bolton, D. (2008). The epistemology of randomized, controlled trials and application in psychiatry. *Philosophy, Psychiatry & Psychology*, 15(2), 159–165.
54. Szasz, T. (1961). *The myth of mental illness*. New York: Harper and Row.
55. HM Government. (2011). No health without mental health: A cross-government mental health outcomes strategy for people of all ages. London, UK: Department of Health. <https://www.gov.uk/government/publications/no-health-without-mental-health-a-cross-government-mental-health-outcomes-strategy-for-people-of-all-ages-a-call-to-action>.

56. NHS England, & NHS Improvement. (2018). The Improving Access to Psychological Therapies (IAPT) pathway for people with long-term physical health conditions and medically unexplained symptoms. <https://www.england.nhs.uk/publication/the-improving-access-to-psychological-therapies-iapt-pathway-for-people-with-long-term-physical-health-conditions-and-medically-unexplained-symptoms/>.
57. Beauchamp, T. L., & Childress, J. F. (2012). *Principles of biomedical ethics* (7th ed.). New York: Oxford University Press.

**Open Access** This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

