CAUSAL EXPLANATION IN PSYCHIATRY

Tuomas K. Pernu

Department of Philosophy, King's College London
&
Molecular and Integrative Biosciences Research Programme, Faculty of Biological and Environmental Sciences, University of Helsinki

www.tuomaspernu.london

Please cite as:
Pernu, Tuomas K. (2019). "Causal explanation in psychiatry". In Ş. Tekin & R. Bluhm eds, *The Bloomsbury Companion to Philosophy of Psychiatry*. London: Bloomsbury Academic.

1. Introduction

The central aim of our scientific endeavour is to give us an accurate picture of the causal structure of the world. Having a valid and precise understanding of real causal relationships lends them to manipulation and control, which is something that is useful across the disciplines. This aspiration is particularly strong in the health sciences, which aim to give us understanding of the causes and mechanisms of diseases in order to enable us to make efficient clinical and preventative interventions. Among the most important questions of philosophy of medicine should therefore be these: what is "causation", what is "causal explanation" and what is "causal efficacy"?

Although causal explanation occupies a central place in all the health sciences, psychiatry is special in that it is affected by the mind-body problem and the issue of mental causation, the question of how, or whether, the mental and the physical can interact with each other. This chapter provides tools with which to analyse these issues in psychiatry and discusses some ways of tackling these problems in the light of some recent developments in the philosophy of science. Although these problems are not easy, and they are particularly grave in the context of psychiatry, recent discussion has brought about some significant advances which have the potential to enhance our understanding of the scientific identity of psychiatry.

The mind-body problem materialises in psychiatric context as a tension between psychological and physiological ways of explaining mental disorders. Your stance on the issue of how the mental and the physical are related, and whether they are in a causal interaction, therefore bears directly on what sort of psychiatric research and what kinds of clinical interventions you are bound to favour. The following will outline the basic philosophical elements of this problem. It is shown how having a solid, philosophically and scientifically backed up account of causation can help one to reconcile the intuition that mental states can function as causes and effects with the idea that reality is thoroughly physical. However, a problem for this view is also sketched. Holding on to the idea of genuine (autonomous) mental causation requires one to assume that mental states cannot be identified with neural (or bodily) states. Although such an assumption can find support from current psychiatry, there are also reasons for scepticism, as shown in the end.

2. Causation in science and medicine

2.1 What is scientific explanation? Let us take a few steps back and look at the scientific enterprise from a distance. What is it that we are trying to do in science? What is the ultimate goal? There are many perfectly good and enlightening answers to this question, of course, but at the same time it is clear that one answer is particularly pertinent: we are trying to get an accurate picture of the causal structure of the world. Why? Because having the correct understanding of the actual causal relations obtaining in the world enables us to implement effective strategies to control and manipulate the world according to our needs and desires. If there is a pragmatic interest to science, as there surely is, then this must be it.

Naturally there is much more to science than searching for causal explanations. For example, it is not at all clear – or at least it would be highly controversial to claim – that such abstract fields as philosophy, logic, mathematics or theoretical physics are in any way engaged in an enquiry aiming at giving us an understanding of the causal structure of the actual world. However, at the same time it is clear that a large part, if not all, of the sciences we deem "empirical" are doing exactly that, from chemistry to psychology, and from biology to economics. Moreover, medicine is a field of enquiry that fits this model particularly well, and in several, interconnected ways.

2.2 Causation is central to medicine: diagnosis, prevention, treatment. Aetiology is a central notion in medicine. Aetiology, as it is typically understood, is the study of the causes or origins of a disease, disorder, or a condition that is thought to require medical attention. If you give it a moment's thought, and decipher the semantics of the term in your head, you'll see that it bears two different meanings or functions in medicine. Firstly, aetiology is an essential element of diagnostics: knowledge of where and how a particular condition has originated plays a crucial role in determining the identity of the condition. What's central to medicine, and what's of particular concern in psychiatry, is the demarcation between the pathological and the (physiologically/psychologically) "normal". Pathology, in turn, is the study of causes and origins of a disease or a disorder – of the external source (pathogens) of abnormal, dysfunctional state of an organ or organism. It should be clear, therefore, that having reliable information on the origins of the condition is diagnostically useful. Moreover, such information is useful beyond diagnosing particular cases, for secondly, and maybe more importantly, medicine is not concerned only with studying diseases, but also, and primarily, with preventing diseases and other conditions that we deem harmful. Naturally, the more accurate information we have on the causes of a particular condition, the more successful we'll be in preventing that condition from arising.

But there is another, perhaps even a more essential way in which causal information becomes entangled with medical practice. Let us use etymology as our guide again: the term "medicine" means the practice of providing attention and care to the sick and injured (Charen 1951). This is what medicine aims at, first and foremost: to alleviate suffering, both physical and mental, and, ultimately, to heal those who are suffering from diseases, disorders, injuries or other conditions that are considered harmful. In other words, medicine aims at specific, rather tangible and concretely useful *effects*, both at the personal and at the populational level. And this in turn makes the search for effective treatments and practices the main focus of medicine. There is thus a constant need to determine the effective interventions *vis-à-vis* a particular condition, and to set these carefully apart from those interventions that are not effective (or not *as* effective, *vis-à-vis* this particular condition). At one extreme this core nature of medicine materialises as a fight against quackery and "alternative medicine" – as a fight against ineffective or even harmful interventions, advertised as effective.

From this concern arises also the more general interest in the constant assessing of the efficacy of medical practices and in developing better tools for the accurate assessment of

such practices – that is, the very idea of "evidence-based medicine" – and, consequently, the ever-growing interest in the placebo effect, the need to separate the "really effective" medical interventions from the merely apparently effective ones. Medicine, it thus seems, is thoroughly engaged with the idea of causal efficacy.

Since the notion of causation is so central to medicine, among the most important questions of philosophy of medicine should be these: what is "causation", what is "causal explanation" and what is "causal efficacy"? These are important, and old, questions of metaphysics and philosophy of science, and one must be prepared to approach the depth and complexity of the problems associated with them with a humble attitude. At the same time, however, it would be wrong to let oneself fall into despair and conclude that such questions are too profound and difficult to merit any systematic attention. This would be not just unfortunate but ill-founded, for many recent developments in metaphysics and philosophy of science have in fact rendered these issues systematically tractable. Philosophy has progressed, and analysing the issue of causal explanation in psychiatry is actually a particularly enlightening way to demonstrate this. Having a rigorous understanding of some of the basic ideas of recent philosophy of science and philosophy of mind is quite concretely useful in addressing many foundational questions of psychiatry.

3. What is causation?

3.1 Causation: dependency vs production. Although there are a number of well-developed accounts of causation in current philosophy (agential, counterfactual, interventionist, physical process, probabilistic, regulative *etc*.), one can make one very useful taxonomical distinction, a distinction that is particularly useful in the current context. On the one hand, there are accounts that stress the idea that causation is a matter of interdependence of events or variables. On the other hand, there are accounts that stress the idea that causation is a matter of physical process or production (Hall 2004). The first camp includes, most prominently, the counterfactual (Lewis 1973), interventionist (Woodward 2003), probabilistic (Eells 1991; Suppes 1970; Williamson 2004) and structural equation or causal modelling (Halpern 2016; Pearl 2000; Spirtes *& al*. 2000) accounts. The second camp includes the causal line (Russell 1948), conserved quantity (Dowe 1992, 2000; Salmon 1997), energy transference (Castañeda

1980; Fair 1979), mark transmission (Salmon 1984), physical force (Bigelow *& al*. 1988; Bigelow & Pargetter 1990) and trope persistence (Ehring 1997, 2003; Kistler 1998, 2006) accounts. There is significant overlap between these views, of course, but this overlap is typically between accounts falling into the same category. Very rarely, if ever, do insights from one account seep through the boundary into another account in the competing camp.

This distinction is particularly relevant in the context of psychiatry for the tension between psychological and physiological explanations can be seen to be rooted in this distinction, or something closely akin to it. Although there are various subtle differences in the theories, there is one central point of departure to draw attention to: the different attitudes the two separate types of theories of causation hold on the issue of causal locality (*cf*. Dowe 2004; Schaffer 2004). We will need to rely on a very crude and intuitive notion of locality here, but the basic idea should be relatively clear. What the dependency accounts require of a proper causal relation is merely that some parts of the world (or representations of them) are dependent on each other in a way specified by the theory (counterfactually, probabilistically *etc*.). What the production accounts require, on the contrary, is that these parts of the world (and not their representations) are physically connected to each other (or that the causal dependency can always be traced back to such a physical, concrete connection). In other words, what the latter accounts require is that there is a continuous, unbroken chain of events from causes to effects; their differences lie in the ways they specify what this continuous chain of events really amounts to.

3.2 Which view of causation is more "scientific"? The physical production view of causation could initially seem more solid and scientific, but such an appearance is deceiving. Firstly, although it is intuitively credible to think that genuine causal relations are concrete, local interactions – this is something that the physicalistic, scientific world view would seem to suggest – a broad range of intuitively clear cases of causal interaction do not trade on such ideas, at least not in any obvious way. For example, it seems perfectly right to say that the macro-economical decline in the 1930s, the Great Depression, caused a number of suicides across Europe and North America. It is not hard to think other examples of intuitively cogent causal claims that we find impossible to translate into the language of physical pushing and pulling. Moreover, omissions and absences can typically function as perfectly good causes and effects, but such things are "nowhere", *per definitionem*, and hence would seem to be ill-

suited to figure in any local physical interactions. For example, it seems perfectly right to say that the patient's failure to take her medication caused her psychotic episode. If the failure, the absence of medication, is functioning as a cause in this scenario, it clearly cannot act as a local producer of the effect.

Secondly, and more importantly, the dependency view reflects better how causal reasoning occurs in a variety of scientific disciplines than the physical production view. Science works through, in most typical cases, by analysing datasets to identify stable and recurring connections between different data points. What one tries to do, in other words, is to spot genuine causal relationships from empirical data by using various statistical methods. The pretheoretic causal framework applied in this sort of analysis relies on statistical (counterfactual, probabilistic) dependencies, not on concrete physical production. Looking at the actual scientific practice, it would therefore be more appropriate to hold the dependency view as the "more scientific" of the two.

However, there is a subtle connection between the two views that often fails to get as much attention as it deserves. The main philosophical problem here is that it seems that one needs to supplement the purely statistical analysis with other, more concrete information in order to reach reliable causal conclusions. Statistical analysis will of course deliver only statistical results, but causation is naturally something different to statistical correlation. In the health sciences in particular, simple statistical evidence is rarely taken to be enough to ground causal claims. To reach generally accepted causal conclusions, the statistical evidence needs to be supplemented with mechanistic understanding (*cf*. Russo & Williamson 2007). Smoking, for example, is generally thought to cause lung cancer not just because smoking and lung cancer are statistically linked, but because we know what the ingredients of cigarettes are, and we know how consuming them is mechanistically linked to changes at the cellular level that give rise to cancer. And now, when you start to analyse this reasoning, you easily fall back into viewing causation as some sort of local relationship of physical production. This tension, or an interplay, if you wish to see the connection in a more positive light, between these two views on causation can be thought to be particularly tangible in psychiatry where the issue of how mental disorders and their physiological underpinnings are connected is constantly and concretely present.

3.3 The interventionist account of causation. This caveat notwithstanding, let us now have a closer look at one well-defined account of causation. To make the discussion as precise as possible, the following will rely on an interventionist account of causation, a paradigmatic representative of the dependency view. There are many reasons to adopt this account. First, the view is precisely defined and widely studied in recent philosophy of science, and one could say that it has become the dominating view on causation in current philosophy. Second, the account is closely connected to scientific practice, and it suits analysing causal claims in the health sciences particularly well. Third, it has received attention in recent philosophy of psychiatry, and it is claimed that new solutions have been found for a number of fundamental problems in psychiatry with the help of this sort of understanding of causation (*e.g.* Campbell 2008a-b, 2009; Kendler 2011; Kendler & Campbell 2009; Woodward 2008).

The interventionist account of causation integrates many insights from different accounts of the dependency view, the agency, counterfactual and structural equation accounts in particular. According to interventionism, causal claims are claims about results of hypothetical interventions in counterfactual scenarios (*e.g.* Halpern 2016; Pearl 2000; Woodward 2003). There are two basic elements to this account. First, you need to specify a set of variables (*e.g.* {*C*, *E*}) that constitutes the domain of entities or events under scrutiny. Second, you define a set of structural equations specifying the various dependencies of the variables in the domain (*e.g.* (if $C = 1$, then $E = 1$) and (if $C = 0$, then $E = 0$)). Causal relations are, therefore, simply patterns of correlations among the values of the variables under hypothetical changes in them. More precisely: a variable *C* is a cause of a variable *E* (in the given domain) just in case there is an intervention on the value of *C* that will result in a change in the value of *E*.

There are of course various technical details to this account, but this rough idea should be enough for outlining the basic issues we are faced with in psychiatry regarding causal explanation. However, two philosophical issues are worth mentioning. First, interventionism offers a nonreductionistic analysis of causation. Note that typical analyses are reductionistic in that they analyse the notion of causation fully in non-causal terms: for example the counterfactual account analyses the causal relation to counterfactual dependencies, the probabilistic to probabilistic dependencies, the transference account to transference of energy, and so on. Intervention, however, is clearly a causal notion, and it is used to define what causation is. Secondly, interventionism is anthropocentric, or at least has a strong

anthropocentric element to it: causal claims make sense only in contexts where the relevant interventions can be carried out. Some might object that this leaves a number of objective causal processes outside of the analysis – how should we account for the cosmic effects of black holes, for example? – and argue that we therefore should not accept interventionism as a complete, or final analysis of causation. Although it is important to be aware of both of these critical issues, it can be assumed that the analysis is still useful, in the current context at least, as the following will demonstrate.

4. Causation and "levels of reality"

4.1 Psychiatry and the mind-body problem. Understanding causal explanation is central to understanding how medicine works. In psychiatry, however, we are faced with the issue of causal explanation in a setting that is particularly challenging. Although all medicine is complex – in the sense of being multifactorial and multilayered – psychiatry occupies its own level of complexity: psychiatry is trying to understand the mental realm, and to do so by navigating in-between the mental and the physical views on reality. In other words, basically all of psychiatry is thoroughly entangled with the mind-body problem, and the issue of how (or whether) the mental and physical can interact.

Deciphering the relationship of the mental and the physical is notoriously difficult. However, this is yet another issue where actual philosophical progress has been made, and we are in a position to formulate the problem, or at least some relevant parts of it, more precisely than ever, and therefore also able to reach precise results with the potential for concrete applications. Having a proper understanding of these developments is crucial to understanding the foundations of psychiatry.

What is "mental" then, as opposed to "physical"? What is the distinction? Although there are a number of things to draw attention to (*cf*. Pernu 2017), the following two characteristics are particularly pertinent in this context: first, our mental states are meaningful, that is, they refer to things and events outside of themselves and outside of the brains that ground them; second, our mental states, sensations in particular, are accompanied by specific phenomenal, subjective content, a feeling of what it is to have that particular experience. In other words, intentionality and subjective consciousness are the central

characteristics of our mental lives. Perhaps nothing is more real to us than the phenomenal and semantically meaningful content of our mental states. Yet, we find it difficult, if not utterly impossible, to explain them in terms we otherwise hold ultimately real – in terms of the objective and mechanistic sciences of physics and biology.

4.2 Nonreductive physicalism and the idea that the mental is multiply realised by the physical. Let us now simply take for granted that speaking in psychological terms is natural and useful for us, and that replacing this way of speaking with a thoroughly physicalistic parlance is not feasible for us. But the question still remains: how exactly should we conceive of the mental as distinct from the physical? How should the difference between the two be understood in order for us to make sense of the connection that we also perceive to be there between the two? The typical way to think about the relationship in current philosophy is to conceive of the mental as being fully dependent on, yet distinct from the physical. How to have one's cake and eat it too? Hold that the mental is always physically realised – that whatever constitutes the physical basis of a particular mental state determines the occurrence of that mental state in its entirety – but that the mental is multiply realised – that each mental state could have had an alternative physical basis. This is the basic idea of nonreductive physicalism, the dominant view in current philosophy of mind.

It is not difficult to appreciate this view, and it is easy to see how it can be at home in psychiatry. It's monistic: the world is ultimately physical, with no separate realms or entities. But it leaves room for the autonomy of psychology: the mental is dependent, but not identical or reducible to the physical. No wonder the view is popular; it seems to offer something for everyone. How, then, should the core idea, the thesis of multiple realisability, be understood? According to this view the same mental states – or in principle all higher-level, functional states – could, both in principle and in practice, appear in various different material constitutions. Different people share the same thoughts even if their brains are not identical, different species share similar mental functions even if they are biologically radically different, and computers and robots can behave intelligently, and they are designed to resemble us, yet their material constitution is completely different from ours. The idea of multiple realisation seems very natural to us.

At the heart of this thesis is the idea that (psychological) functions are implementation-independent. In computing, you can implement the same software in different hardware, as is

often done, and in principle you could build powerful computers out of sticks and rocks, or so the thinking goes. It is indeed quite attractive to think of psychiatry as a science focusing on the software of the mind, and neurology on the hardware. It is in many ways an illuminating and apt analogy. Both ways of viewing the system (the mind-brain system) are right, there exist efficient interventions at both levels, and there is a neat division of labour where the usefulness of both types of engineer is recognised. You need a functional hardware to run the given software. But often the malfunctions that appear are malfunctions of the software, not of the hardware.

Let us now suppose that we can in this way account for the intuitive idea that there are various "levels of reality" or "levels or organisation" out there. Molecules are made of atoms, organelles out of molecules, cells out of organelles, organs out of cells, organisms out of organs, and so on. Minds, on this view, are simply another level of reality arising from the right sort of biological organisation – brains and central nervous system. Note that this way of construing the levels of reality as a nested functional hierarchy incorporates the idea of multiple realisability: entities at one level retain their identity even under radical changes at the lower-levels – like organisms retain their identities even though they go through constant changes at the cellular level. Thus, we can hold onto the idea that there are distinct, higher-levels to reality, the mental realm among them, that are dependent on the lower-levels that realise them, and are nothing "over and above" them.

4.3 Causation in nonreductive physicalism. If we accept this view, how, then, should we perceive causation? What sort of causal interactions does this view allow? The received view, and the main thrust of nonreductive physicalism, is the idea that the higher-level entities and events are genuinely causally efficacious. In fact, this is the reason for holding them "real": higher-level entities and events have irreducible causal powers, and therefore we are committed to granting them autonomous existence. At the same time, however, the grounding, physical level is thought to be causally complete. In other words, nonreductive physicalism holds that each physical level event that has a cause, has a sufficient, complete physical cause. In this way, we do not need any higher-level information to account for the events occurring at the fundamental physical level, but at the same time, the view maintains, no information confined to the fundamental physical level could be sufficient to account for the causal relationships obtaining at the higher levels.

Let *M* and *M** now be variables standing for higher, mental-level events, and let N_1 and N_2, and N^*_1 and N^*_2 be variables standing for their respective lower, neural-level realisers. Figure 1, the iconic diagram of nonreductive physicalism (*cf.* Fodor 1974), illustrates how this view perceives the relationships between these variables.

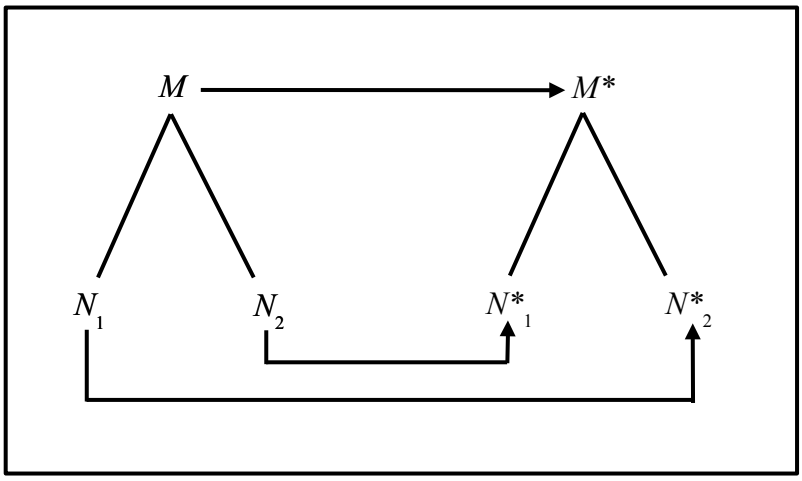

Figure 1: the iconic diagram of nonreductive physicalism. (Adapted from Fodor (1974).)

This view seems to fit well with actual psychiatric practice: no-one doubts that mental events, and mental disorders, are neurally grounded, and that the neurophysiological realm forms a complete system, but at the same time it is the psycho-social interactions that function as typical determinants of the phenomena that call for psychiatric attention.

5. The fragility of psychiatric kinds

5.1 It is wrong to pit social, psychological, and neural interventions against each other (even if one holds to thoroughly physicalistic metaphysics). Although the idea that mental states are neurally grounded should be uncontroversial, the direct consequences of this idea are easily underappreciated. First, it should be obvious that changes at the different levels of analysis, mental and neural, are correlated. Therefore, you cannot support claims on the primacy or autonomous existence of one level by relying solely on evidence showing that changes on this level result in changes on the other level (*cf.* Pernu 2011). For example, noting that certain psychological features have a specific cortical basis does not licence the conclusion that those cortical features are causes of those psychological features (and hence

that those psychological features are somehow less real). Or, conversely, the fact that psychological interventions (psychotherapy) result in changes at the neural level (cortical level), does not say anything about the efficacy of such interventions: we already knew that psychological interventions result in psychological changes, and we knew that mental states are neurally based, so we also know that the resulting psychological changes will always manifest as neural changes.

Another direct consequence of the sketched metaphysical view deserves to be highlighted: we can take it for granted that psychiatric disorders are typically multifactorial, which suggests that various different interventions can be effective, to varying degrees. It is therefore wrong to pit social, psychological, and neural interventions against each other. For example, consider the fact that there is a strong positive correlation between experiencing abuse in childhood and suffering from various mental disorders (depression, anxiety, *etc*.) in adulthood. What sort of interventions should we apply to improve the situation? There is clearly no single right answer; it depends on what sort of effects you are aiming to produce, and at which point in the process you are able, or willing, to act. If you are a clinician, and you are faced with a patient with a unique history and personality, your interest is in finding an effective way of alleviating the current symptoms of this particular person. One effective way of doing that could be to apply neural-level interventions, *i*.*e*. pharmacological treatment. But it would be wrong to conclude from this that mental disorders originating from childhood abuse are nothing but brain disorders, and that pharmacological interventions are the only correct way to treat them. Obviously, if you are treating an adult patient with a condition that has resulted from childhood events, you are unable to intervene on the ultimate causes of the condition simply because you are unable to intervene on events in the past. However, if you are not faced with a particular patient, but the general issue of how to prevent these types of mental disorders from occurring, if, in other words, your effect-variable is a future oriented population-level one, the recommended intervention should look totally different: you should intervene on the societal conditions that give rise to abusive behaviour. All this should be rather obvious. Nevertheless, all too often psychiatric interventions targeting different levels of organisation are treated as mutually exclusive. Neural-level interventions do not have to be the only appropriate interventions even if all mental states have, by necessity, a neural basis.

5.2 Are mental states really multiply realised? Since the aetiology of mental disorders is multifactorial, and the same types of disorders can rise through radically difference routes, it is natural to assume that mental disorders are multiply realised at the neural level. However, the idea of multiple realisation cannot be taken for granted, especially in the context of current psychiatry.

There are a number of reasons to be critical of the thesis of multiple realisation (*e.g.* Bechtel & Mundale 1999; Polger & Shapiro 2016; Shapiro 2000). The core of the criticism can be phrased in abstract metaphysical terms: if we deem an entity multiply realised, then it would seem that its realisers would have to differ in their causal profiles; but if the realisers differ in their causal profiles, it is not clear why they should be treated as instances of the same multiply realised entity. Is it credible to assume that mental states can be just relevantly similar enough but also just relevantly different enough to be qualified as multiply realisable? A moment's reflection should make one suspect that such a position is hard to hold.

There is an internal tension in the multiple realisability thesis, and that tension can be discharged in two different ways. On the one hand, the assumed multiply realised higher-level entity can split into separate entities each aligned with their realisers. On the other hand, the realisers that were assumed to be distinct can merge into a single realiser. Both of these possibilities dissolve the assumed multiple realisation into identity (symmetrical dependence) between the higher and lower-level entities (which would prompt one to be critical towards the whole stratified picture of reality). Figure 2 illustrates the situation.

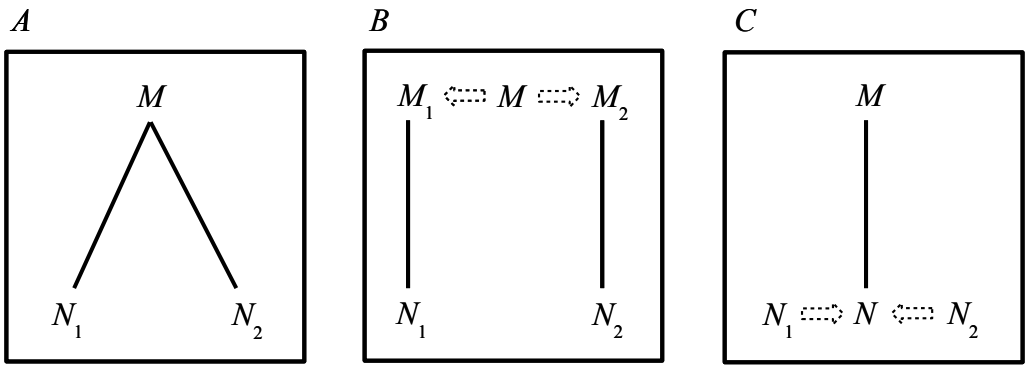

Figure 2: figure 2A represents the multiple realisation hypothesis, figure 2B represent kind splitting, and figure 2C realiser unification or merging. (From Pernu (forthcoming).)

If the problem would be merely conceptual or metaphysical in nature, the issue could be put aside and ignored. That is not the case, however. On the contrary: many of the fundamental issues we currently face in psychiatry are directly linked to the question of

whether the thesis of multiple realisability applies to mental disorders. Moreover, this issue isn't merely theoretical. Different solutions to the issue have direct repercussions for how we position psychiatry among sciences, for what lines of research to undertake, and for what sort of clinical interventions to favour.

5.3 Two examples: schizophrenia and addiction. Consider the following example. In autumn 2017 the British Psychological Society published an updated report on *Understanding Psychosis and Schizophrenia* (Cooke *& al*. 2017). This created a debate – not unexpectedly – on whether the report had taken all the relevant symptomatic, aetiological and pharmacological issues into account. What this debate prompted is an editorial in *The Lancet Psychiatry* (Editors 2018), urging us "to think of schizophrenia as an unmapped, ill-defined area, perhaps as an iceberg" (p. 1). And as science progresses, the editorial predicted, we would witness how "minute but significant parts of this iceberg will break off" (Editors 2018, p. 1). Now, whether or not you subscribe to this train of thought, you have to appreciate its foundations. The idea is that we don't know enough of the causes and effects of schizophrenia – or this thing we call "schizophrenia" – to properly understand its essence. And further, once we get there – once we have that understanding, or something close to it – we realise that there never really was such a thing, that there never was a unified, coherent mental disorder we should deem "schizophrenia". Our proper (concrete, physical, that is) understanding of the phenomenon has actually broken down and thus erased the phenomenon itself. There is no "schizophrenia", but only these separate chips of it, in the vein of Figure 2B.

Consider another, equally central issue in psychiatry: the problem of addiction. There is a debate about whether addiction should be construed as a choice or whether it should be construed as a (neurological) disease (*e.g*. Goldstein & Volkow 2011; Heyman 2009; Leshner 1997; Lewis 2015, 2017; Robinson & Berridge 2000; Szalavitz 2016; Volkow *& al*. 2016). But maybe this juxtaposition is ill-founded? A number of studies have indicated that only a minority of subjects develop a "pathological" substance addiction (*e.g*. Ahmed 2010; Cantin *& al*. 2009; Deroche-Gamonet *& al*. 2004; Dutra *& al*. 2008; Robins *& al*. 1974). What this suggests, then, is that there might be different types of "addiction" – that addiction is not one coherent psychological or behavioural kind. And these different types of addiction, in turn, would call for different types of interventions. So, to be more analytical, first, we observe that distinct clusters of subjects respond in distinct ways to addictive cues – distinct clusters of

testees have distinct causal profiles *vis-à-vis* addictive scenarios. And second, we observe (or we should observe) that different types of interventions are efficacious, depending on which cluster the patient belongs to. This, again, should prompt us to split "addiction" into (at least two) different types, in accordance with their causal profiles. As a result, "addiction" would not be genuinely multiply realised, but would encompass different things, each in alignment with their neural bases, in the vein of Figure 2B. Of course we might still have pragmatic reasons to keep using the one concept, and maybe even to treat the different types of addiction in the same institutions, but we should not let such social conventions mislead us into thinking that these different things are fundamentally the same.

5.4 The increasing pressure to dissolve the apparent multiple realisation of mental disorders: the RDoC framework. To further motivate the thesis that psychiatry is in a particularly fragile state – "fragile" in the sense that its current nosological practices are under pressure to become more fine-grained – consider the introduction of the Research Domain Criteria (RDoC) framework into psychiatric research (Cuthbert 2014; Cuthbert & Kozak 2013; First 2012; Insel *& al*. 2010). The traditional Diagnostic and Statistical Manual of Mental Disorders (DSM) framework is based on symptomatic classification of disorders. This way of classifying mental disorders is prone to produce invalid results, the critics claim, for clusters of symptoms are not robust enough to provide us accurate and stable information about the disorders. That is why we should look into the various ways mental disorders are actually (biologically) realised, and aim to classify them in a more accurate manner.

It is not difficult to appreciate the attractiveness of the RDoC framework. Although the framework is aimed primarily at improving research on mental disorders, it is clear that there is a heavy clinical thrust behind the initiative. In fact, one argument against the DSM, and in favour of the RDoC, has targeted the influence that the DSM has on drug development, and how that leads to suboptimal results due to the DSM based categories lacking biological validity (*e.g*. First 2012; Hyman 2010). Having more accurate, valid classifications of mental disorders provides us with means to reach more accurate diagnosis, and, consequently, to apply more accurate and effective medical interventions. Having a better understanding of the biological constitution of a particular disorder makes us able to make better predictions of how the disorder behaves in different situations and under different interventions – in the

same way that having correct understanding of the phylogenetic positioning of an organism allows us to make reliable predictions about the traits and behaviour of that organism.

Whether this aim can be achieved depends on whether mental disorders are genuinely multiply realised – whether, at least in some important cases, we are faced in reality with the figure 2A rather than the figure 2B (*cf. e.g.* Hoffman & Zachar 2017; Parnas 2014). If that is the case, then having appropriate information of the biological realisers of these mental disorders will not allow us to make optimally accurate and stable predictions. The jury is still out, and we will need more both conceptual and empirical research to reach the right verdict. However, it is clear that many indicators point to the conclusion that the idea of multiple realisation of mental disorders cannot be taken for granted, and that many disorders now treated as homogeneous will split into separate disorders.

6. Concluding remarks

We are living exciting times in psychiatry, and especially in research on the foundational issues. There have been significant advances in the philosophy of science, especially on causation and on issues related to reduction and multi-level explanation. We can see more clearly than ever how our views on causation bear on how we think the mind and body are connected and this has potential to give us better understanding also on how the apparent tension between psychological and neural interventions could be discharged. At the same time, empirical research has developed more accurate tools, and more precise data has been accumulated on the biological underpinnings of mental disorders. This will lead, the hope is, to a more valid psychiatric nosology, with recognised disorders having more homogeneous causal profiles, which lends the disorders to more precise and efficient manipulation and control – to more precise and effective treatments. This hope might of course turn out to be ill-founded; we might have to admit that mental disorders are genuinely multiply realised at the physiological level, and that it is psychological, rather than pharmacological interventions, that are most effective in typical clinical cases. But whichever way we go the result will depend on our philosophical understanding on causation and multi-level explanation.

It is therefore not a stretch to claim that we are starting to have the elements in place which will enable us to gain significant advances in understanding the nature of mental disorders, if only the elements are put together the right way. To make that step, and to make that decisive advancement, what is needed is a more intimate collaboration between the conceptual and the empirical strains of research. Intellectual silos should therefore be abandoned, and more resources should be allocated to genuinely interdisciplinary research.

Acknowledgements. I would like to thank Dr Robyn Bluhm, Mr Peter Cave, Dr Nadine Elzein, Dr Outi Mantere, Dr Hane Maung, Mr Ivo Storrie, Dr Şerife Tekin, the participants of the Philosophy & Medicine Reading Group at King's College London in the winter/spring of 2017, and the participants of the Philosophy of Psychiatry Seminar at King's College London on 22 January 2018 for helpful discussions on the topic and comments on various versions of this paper. This work has been financially supported by The Finnish Academy of Science and Letters, the Emil Aaltonen Foundation and the Waldemar von Frenckell foundation.

References

Ahmed, Serge H. (2010). "Validation crisis in animal models of drug addiction: beyond non-disordered drug use toward drug addiction". *Neuroscience & Biobehavioral Reviews* 35, p. 172-184.

Bechtel, William & Mundale, Jennifer (1999). "Multiple realizability revisited: linking cognitive and neural states". *Philosophy of Science* 66, p. 175-207.

Bigelow, John; Ellis, Brian & Pargetter, Robert (1988). "Forces". *Philosophy of Science* 55, p. 614-630.

Bigelow, John & Pargetter, Robert (1990). *Science and Necessity*. Cambridge: Cambridge University Press.

Campbell, John (2008a). "Causation in psychiatry". In K. S. Kendler & J. Parnas eds, *Philosophical Issues in Psychiatry: Explanation, Phenomenology, and Nosology*. Baltimore: Johns Hopkins University Press.

Campbell, John (2008b). “Comment: psychological causation without physical causation”. In *Philosophical Issues in Psychiatry: Explanation, Phenomenology, and Nosology*, ed. K. S. Kendler & J. Parnas. Baltimore: The Johns Hopkins University Press.

Campbell, John (2009). “What does rationality have to do with psychological causation? Propositional attitudes as mechanisms and as control variables”. In M. Broome & Lisa Bortolotti eds, *Psychiatry as Cognitive Neuroscience: Philosophical perspective*s. Oxford: Oxford University Press.

Castañeda, Héctor-Neri (1980). “Causes, energy and constant conjunctions”. In P. van Inwagen ed., *Time and Cause*. Dordrecht: Reidel.

Cantin, Lauriane; Lenoir, Magalie; Dubreucq, Sarah; Serre, Fuschia; Vouillac, Caroline & Ahmed, Serge H. (2010). “Cocaine is low on the value ladder of rats: possible evidence for resilience to addiction”. *PLoS ONE 5*: e11592.

Charen, Thelma (1951). “The etymology of medicine”. *Bulletin of the Medical Library Association* 39, p. 216-221.

Cooke, Anne *& al*. (2017). *Understanding Psychosis and Schizophrenia*. British Psychological Society, Division of Clinical Psychology.

Cuthbert, Bruce N. (2014). “The RDoC framework: facilitating transition from ICD/DSM to dimensional approaches that integrate neuroscience and psychopathology”. *World Psychiatry* 13, p. 28-35.

Cuthbert, Bruce N. & Kozak, Michael J. (2013). “Constructing constructs for psychopathology: the NIMH Research Domain Criteria”. *Journal of Abnormal Psychology* 122, p. 928-937.

Deroche-Gamonet, Véronique; Belin, David & Piazza, Pier V. (2004). "Evidence for addiction-like behavior in the rat". *Science* 305, p. 1014-1017.

Dowe, Phil (1992). "Wesley Salmon's process theory of causality and the conserved quantity theory". *Philosophy of Science* 59, p. 195-216.

Dowe, Phil (2000). *Physical Causation*. New York: Cambridge University Press.

Dowe, Phil (2004). "Why preventers and omissions are not causes". In C. Hitchcock ed., *Contemporary Debates in Philosophy of Science*. Oxford: Blackwell.

Dutra, Lissa; Stathopoulou, Georgia; Basden, Shawnee L.; Leyro, Teresa M.; Powers, Mark B. & Otto, Michael W. (2008). "A meta-analytic review of psychosocial interventions for substance use disorders". *American Journal of Psychiatry* 165, p. 179-187.

Editors (2018). "Breaking the ice". *The Lancet Psychiatry* 5, p. 1.

Eells, Ellery (1991). *Probabilistic Causality*. Cambridge: Cambridge University Press.

Ehring, Douglas (1997). *Causation and Persistence*. Oxford: Oxford University Press.

Ehring, Douglas (2003). "Physical Causation". *Mind* 112, p. 529-533.

Fair, David (1979). "Causation and the flow of energy". *Erkenntnis* 14, p. 219-250.

First, Michael B. (2012). "The National Institute of Mental Health Research Domain Criteria (RDoC) project: moving towards a neuroscience-based diagnostic classification in psychiatry". In K. S. Kendler & J. Parnas eds, *Philosophical Issues in Psychiatry II: Nosology*. Oxford: Oxford University Press.

Fodor, Jerry A. (1974). "Special sciences (or: the disunity of science as a working hypothesis)". *Synthese* 28, p. 97-115.

Goldstein, Rita Z. & Volkow, Nora D. (2011). “Dysfunction of the prefrontal cortex in addiction: neuroimaging findings and clinical implications”. *Nature Reviews Neuroscience* 12, p. 652-669.

Hall, Ned (2004). ”Two concepts of causation”. In J. Collins, N. Hall & L. Paul eds *Causation and Counterfactuals*. Cambridge, MA: The MIT Press.

Halpern, Joseph Y. (2016). *Actual Causality*. Cambridge, MA: MIT Press.

Heyman, Gene M. (2009). *Addiction: A Disorder of Choice*. Cambridge, MA: Harvard University Press.

Hoffman, Ginger A. & Zachar, Peter (2017). “RDoC′s metaphysical assumptions: problems and promises”. In J. Poland & Şerife Tekin eds, *Extraordinary Science and Psychiatry: Responses to the Crisis in Mental Health Research*. Cambridge MA: MIT Press.

Hyman, Steven E. (2010). “The diagnosis of mental disorders: the problem of reification”. *Annual Review of Clinical Psychology* 6, p. 155-179.

Insel, Thomas; Cuthbert, Bruce; Garvey, Marjorie; Heinssen, Robert; Pine, Daniel S.; Quinn, Kevin; Sanislow, Charles & Wang, Philip (2010). “Research domain criteria (RDoC): toward a new classification framework for research on mental disorders”. *American Journal of Psychiatry* 167, p. 748-751.

Kendler, Kenneth S. (2011). “Causal thinking in psychiatry: a genetic and manipulationist perspective”. In P. E. Shrout, K. M. Keyes & K. Ornstein eds, *Causality and Psychopathology: Finding the Determinants of Disorders and Their Cures*. Oxford: Oxford University Press.

Kendler, Kenneth S. & Campbell, John (2009). “Interventionist causal models in psychiatry: repositioning the mind–body problem”. *Psychological Medicine* 39, p. 881-887.

Kistler, Max (1998). “Reducing causality to transmission”. *Erkenntnis* 48, p. 1-24.

Kistler, Max (2006). *Causation and Laws of Nature*. London: Routledge.

Leshner, Alan I. (1997). “Addiction is a brain disease, and it matters”. *Science* 278, p. 45-47.

Lewis, David K. (1973). “Causation”. *Journal of Philosophy* 70, p. 556-567.

Lewis, Marc D. (2015). *The Biology of Desire. Why Addiction Is not a Disease*. New York: Public Affairs.

Lewis, Marc D. (2017). “Addiction and the brain: development, not disease”. *Neuroethics* 10, p. 7-18.

Parnas, Josef (2014). “The RDoC program: psychiatry without psyche?”. *World Psychiatry* 13, p. 46-47.

Pearl, Judea (2000). *Causality: Models, Reasoning, And Inference*. Cambridge: Cambridge University Press.

Pernu, Tuomas K. (2011). ”Minding matter: how not to argue for the causal efficacy of the mental”. *Reviews in the Neurosciences* 22, p. 483-507.

Pernu, Tuomas K. (2017). “The five marks of the mental”. *Frontiers in Psychology* 8:1084.

Pernu, Tuomas K. (forthcoming). “Mental causation via neuroprosthetics? A critical analysis”. *Synthese*.

Polger, Thomas W. & Shapiro, Lawrence A. (2016). *The Multiple Realization Book*. Oxford: Oxford University Press.

Robins, Lee N.; Davis, Darlene H. & Goodwin, Donald W. (1974). "Drug use by U.S. Army enlisted men in Vietnam: a follow-up on their return home". *American Journal of Epidemiology* 99, p. 235-249.

Robinson, Terry E. & Berridge, Kent C. (2000). "The psychology and neurobiology of addiction: an incentive-sensitization view". *Addiction* 95, p. 91-117.

Russell, Bertrand A. W. (1948). *Human Knowledge: Its Scope and Limits*. London: George Allen & Unwin.

Russo, Federica & Williamson, Jon (2007). "Interpreting causality in the health sciences". *International Studies in the Philosophy of Science* 21, p. 157-170.

Salmon, Wesley (1984). *Scientific Explanation and the Causal Structure of the World*. Princeton: Princeton University Press.

Salmon, Wesley (1997). "Causality and explanation: a reply to two critiques". *Philosophy of Science* 64, p. 461-477.

Schaffer, Jonathan (2004). "Causes need not be physically connected to their effects". In C. Hitchcock ed., *Contemporary Debates in Philosophy of Science*. Oxford: Blackwell.

Shapiro, Lawrence A. (2000). "Multiple realizations". *Journal of Philosophy* 97, p. 635-654.

Szalavitz, Maia (2016). *Unbroken brain: A Revolutionary New Way of Understanding Addiction*. NewYork, NY: St. Martin's Press.

Spirtes, Peter; Glymour, Clark & Scheines, Richard (2000). *Causation, Prediction and Search*. Cambridge, MA: MIT Press.

Suppes, Patrick (1970). *A Probabilistic Theory of Causality*. Amsterdam: North-Holland Publishing Company.

Volkow, Nora D.; Koob, George F. & McLellan, A. Thomas (2016). "Neurobiologic advances from the brain disease model of addiction". *New England Journal of Medicine* 374, p. 363-371.

Williamson, Jon (2004). *Bayesian Nets and Causality: Philosophical and Computational Foundations*. Oxford: Oxford University Press.

Woodward, James (2003). *Making Things Happen: A Theory of Causal Explanation*. New York: Oxford University Press.

Woodward, James (2008). "Cause and explanation in psychiatry: an interventionist perspective". In K. S. Kendler & J. Parnas eds, *Philosophical Issues in Psychiatry: Explanation, Phenomenology, and Nosology*. Baltimore: Johns Hopkins University Press.