VALIDITY AND THE CAUSAL STRUCTURE OF A DISORDER

BY JOHN CAMPBELL

Forthcoming in Kenneth Kendler and Josef Parnas (eds.), ***Philosophical Issues in Psychiatry IV: Psychiatric Nosology*** **(Oxford: Oxford University Press, expected publication 2016)**

VALIDITY AND THE CAUSAL STRUCTURE OF A DISORDER

This paper looks at the problem raised by there being both psychological and biological causes for disorders. I argue that the position is methodologically relatively straightforward so long as we think of causation merely in terms of intervention counterfactuals or probabilistic relations. Once we think in terms of causal processes, however, it becomes much more difficult to interpret the causal structures here. There are mental processes and there are physical processes, but there does not seem to be any way of connecting the two.

I think that this puzzle forms the background to some of the uncertainty we naturally feel about the validity of psychiatric classifications. I think that the problem is that to have the question of the validity of a classification properly posed, we have to have at least a hypothesis about the causal structure of the disorder. Our uncertainty as to causal structures here means that it's difficult to know how to pose the problem of validity in any particular case.

1. The 'Princess Elisabeth' Problem in Psychiatry

There appear to be both environmental and genetic risk factors for many disorders. There has to be an understanding of how those risk factors are transduced into proximal aspects of the individual that generate those disorders. And at the moment, for many of the environmental risk factors, there is only a psychological understanding to be had of their

implications for the individual. Consider, for example, divorce or bereavement, which are risk factors for major depression. There is no neurobiology of divorce, or bereavement. On the other hand, consider the transduction of genetic risk factors for a disorder. These typically have only a biological transduction into proximal causal factors for a disorder. So the general situation is as diagrammed in Figure 1 for the case of major depression.

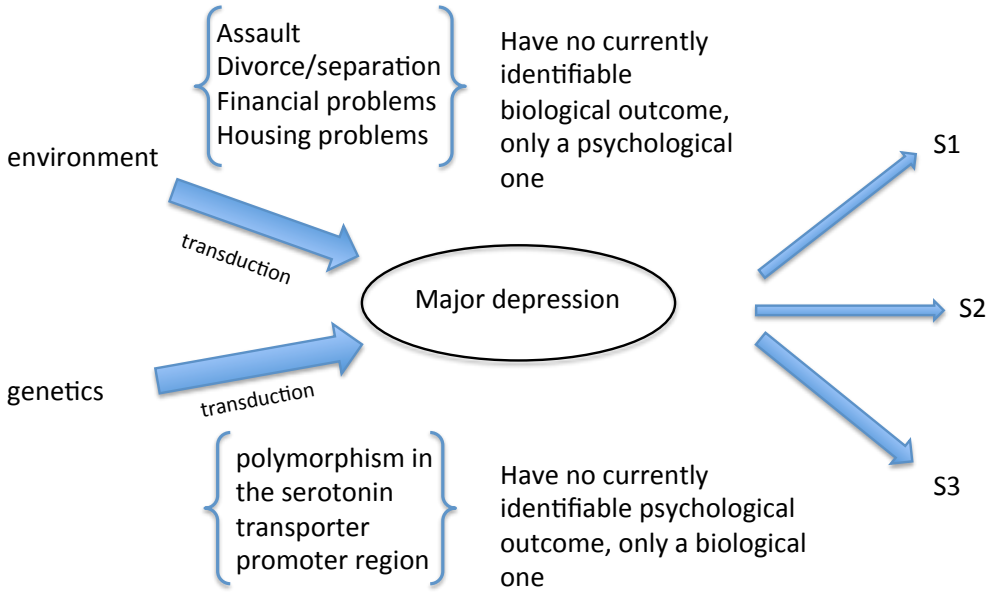

Figure 1. Schematic diagram of one possible causal structure for major depression. Environmental risk factors characteristically have a psychological transduction. Genetic risk factors characteristically have a biological transduction. (With thanks to Ken Kendler.)

At the moment, there's no particular evidence to support the idea that the psychological factors will of themselves coalesce into a progression, comprehensible in purely psychological terms, leading to major depression and its symptoms. The purely biological risk factors may be ineliminable. But nor is there any evidence to support the idea that the psychological interpretations of the environmental risk factors will ultimately be eliminable in favor of purely biological causal factors.

You might argue that the psychological and the purely biological causal structures here are relatively insulated from one another: the psychological factors generate psychological upshots and psychological symptoms, the purely biological factors generate purely biological upshots and purely biological symptoms. But that hardly seems credible as an interpretation. In general, biological factors are relevant to the psychological symptoms of a disorder.

It thus appears that at the moment, understanding the causal structure of a disorder will, in general, require understanding the causal relations between psychological and biological factors. Now there are another two ways in which you might try to get round the need to do this.

(1) You might argue that ultimately, there are only biological causal factors here. All the causation is to be understood at the biological level. At the moment, admittedly, we have to work with psychological factors. But that is only a reflection of our currently limited knowledge of brain biology. In the future, our understanding of the causal structures of disorders will be entirely at the

biological level; talk of psychological risk factors is something we have to do only because of the current limitations on our knowledge.

(2) You might argue that all the biological risk factors for disorders can ultimately be interpreted in psychological terms. Perhaps serotonin imbalances can be interpreted in terms of some psychological construct like psychological resilience, for example, and ultimately the causal structures of disorders can be interpreted entirely in psychological terms.

At the moment, both of these pictures are science fiction. Of course our knowledge of brain biology seems bound to improve. But it's merely speculation to suppose that this will enable us to eliminate talk of psychological risk factors for disorders. It therefore seems worth looking at the difficulties we face if we have to regard psychiatric disorders as having both psychological and biological factors implicated in their etiology.

To focus directly on the theoretical question here, let's simplify the empirical picture. Suppose we look directly at just two risk factors for major depression. Suppose we consider social humiliation as our example of a psychological factor that we take to be not reducible to biology. And suppose we consider genetically based irregularities in serotonin transport as our example of a biological factor that may have no psychological interpretation. Perhaps the example will turn out to be badly chosen, perhaps there will turn out to be a biological reduction of humiliation, and perhaps there will turn out to be a psychological interpretation of serotonin irregularities. As I said, my point is only to investigate the implications of there not being a wholesale biological reduction of the

psychological risk factors for disorders, or it not generally being possible to give a psychological interpretation of the biological risk factors for a disorder; so let's take it for the moment that there's no biological reduction of humiliation and no psychological interpretation of serotonin imbalance. So the situation, for this simple example, is as shown in Figure 2.

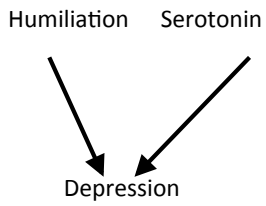

Figure 2. A 'mixed' pair of causal factors for major depression.

Now so long as we think only in terms of correlations, the situation here is relatively unproblematic. Certainly there could, in principle, be correlations between humiliation and depression, and between serotonin imbalance and depression. Indeed the correlations might be such that it's only when we have both humiliation and serotonin imbalance that we have an increased risk for depression. We have, though, to consider the case in which humiliation and serotonin imbalance are joint causes of depression. The problem comes when we try to think through the mechanisms or processes that might be involved in their joint production of the outcome. On the one hand we might suppose that we have a biological reduction of humiliation, so that we can explain the mechanism by which they jointly produce their outcome in entirely biological terms. But we are

trying to explore the consequences of supposing that biological reductions are not always available.

To try to make vivid the theoretical perspective I am suggesting here, consider the position of a social psychologist investigating the consequences of a 'don't ask, don't tell' policy in the military, or the causes of voting behavior – why do people vote for a particular candidate? The beliefs and preferences of the population are likely to figure among the causes of behaviors. But the demand for a biological reduction of those beliefs and preferences is likely to strike our social psychologist as idle. Similarly, you might take the view that psychological factors such as social humiliation figure among the causes of disorders. But you might equally feel that the demand for a biological reduction of those factors is, for the moment at any rate, idle.

On the other hand we might suppose that there is a purely psychological mechanism available here, that there is some psychological interpretation of serotonin imbalance that will allow us to understand, in psychological terms, how depression is generated. But we are exploring the consequences of supposing that psychological interpretations of the biological are not always available. This leaves us, I think, without any understanding at all of how there could be a mechanism by which humiliation and serotonin jointly produce depression.

To explain our ordinary notions of mechanism and process, we first need the distinction between general and singular causation. General causation is exemplified by claims such as 'smoking causes cancer', or 'humiliation causes depression'. It relates variables, such a 'level of smoking' and 'risk of cancer', which may take different values. Singular causation, on the other hand, is exemplified by claims such as 'Sally's smoking

caused her cancer', or 'Bill's humiliation caused his depression'. These claims relate particular things – Sally's smoking, Bill's depression – rather than variables.

Let's start by focusing on singular causation. In many cases, singular causation is a matter of there being a *process* connecting cause and effect. In the physical case, we have a reasonably firm grasp of what constitutes a causal process: at bottom, it's something like what Locke (1690/1975) called 'the transmission of motion by impulse', what Fair (1979) called the 'transfer of energy', or that Dowe (2000) talks of in terms of 'exchange of conserved quantities'. Something like that is the basic idea of a physical process. Biological processes can generally be seen as particularly complex versions of underlying physical processes, conceived in this way. In the psychological case, on the other hand, we think of mental processes as normatively guided sequences of psychological states. So for example, if you're trying to decide whether to apply for a particular job, your 'mental processes' will involve the recruitment of your current knowledge and beliefs, and your objectives, with the goal of trying to make the right decision. Jaspers (1923/1963) famously drew a distinction between (1) 'subjective' and (2) 'objective' psychopathology as follows:

> 1. We sink ourselves into the psychic situation and *understand genetically by* empathy how one psychic event emerges from another.
>
> 2. We find by repeated experience that a number of phenomena are regularly linked together, and on this basis we *explain causally*.
>
> (Jaspers 1923/1963), 301)

Now understanding 'genetically by empathy' is what provides us with our grasp of a mental process – a psychological causal process. It's the basis of our understanding of singular causation in the psychological case. This is how we achieve our knowledge of how 'one psychic event emerges from another', of how Bill's being attacked gives rise to his defensiveness, for example. As Hoerl (2013) argues, the distinction Jaspers is really after here seems to be a distinction between singular causation in the psychological case, which is known about by empathy, and general causation, which can be studied by looking at correlations and experiments across a population.

Now the trouble with thinking of both humiliation and serotonin imbalance as being, in a particular case, the interactive causes of someone's depression, is that we have no idea what a process might be in which these two factors are combined to generate the depression. If we try to think in terms of a purely biological process that generates the outcome, we run into the problem that this leaves out the humiliation. On the other hand if we try to understand the causation in this particular case purely in terms of empathetic understanding, we leave out the serotonin imbalance.

The notion of 'mechanism' stands to general causation somewhat as the notion of 'process' stands to singular causation. If you're told that Sally's smoking caused her cancer, you can ask what the process was by which it did so. If you're told that in general, 'smoking causes cancer', you can ask, 'What's the mechanism?', where what you're looking for is something like: the structure that sustains a particular type of process. So if we're told that humiliation and serotonin combine to generate depression,

we can ask, 'What's the mechanism?'. And then we find that we have no idea at all how we might think of those two factors combining to generate the outcome.

It is very often the case that we can establish general causation without knowing anything about the mechanism by which it works. John Snow (1965) famously demonstrated that the water supply could be a cause of cholera while having only the haziest idea of the mechanism by which contaminated water produces cholera. In a randomized controlled trial of a drug, to find for example whether it prevents breast cancer, if the trial is well-executed then it can provide knowledge of the causal connection even if the experimenters' conjecture as to the mechanism by which the drug is working turns out to be wrong.

Similarly, a randomized controlled trial can demonstrate definitively that the drug is, for example, preventing breast cancer, even though it does not allow one to say, of any one participant in the study, 'The drug was responsible for this person's not getting breast cancer'. The trial works across the population, it doesn't allow you to identify any one case as one in which it was the action of the drug that had this outcome.

This opens the possibility that, looking at trials across a population, we could find that both humiliation and serotonin imbalances were correlated with the outcome of depression. In fact, by exploiting the possibility of 'natural experiments', we could demonstrate a causal connection between these two factors and depression as outcome. We could do this without having the slightest idea how to go about thinking of a 'mechanism' by which these two factors might combine to generate the outcome. And we could do this without having demonstrated, for any particular case, that humiliation and serotonin imbalance caused *this* individual's depression.

Elisabeth, Princess of Bohemia, famously challenged Descartes as to there could be causal interactions between mind and body, given his dualist conception of mind and body as different substances (cf. Mattern (1978) for references and overview). As an undergraduate, this problem always struck me as fake; hadn't Hume (1740/1975) shown that causation was constant conjunction? And of course there can be correlations between mental and bodily states. The problem arises, however, when we think of causation in terms of mechanism and process. We know about mental mechanisms and processes, and we know about physical mechanisms and processes. But we have absolutely no understanding of how there could be mechanisms and processes linking mental and physical. My point in this introductory section has been that the 'Princess Elisabeth' problem is written large in psychiatry. It means that we do not have a clear conception of the causal structures our diagnostic procedures are trying to identify. If we think of causality in terms of mechanism and process, then we can't work with a mixed set of psychological and physical variables in specifying the causal structures we are trying to identify. But we have only optimism supporting the idea that we'll be able to give a purely psychological account of the relevant causal structures; and similarly it is only optimism that supports the idea that we'll be able to eliminate the mental and give a purely physical account of the relevant causal structures.

2. Validity as Reference vs. Assessment of Ways of Finding Out

The idea of validity, as it's usually used in connection with scientific constructs, has two dimensions. One kind of case is when we have a phenomenon, some characteristic of objects such as mass or temperature, and we want to know if a particular way of identifying its presence or measuring it is any good. So, for example, we might all agree that there is such a thing as 'general intelligence' in humans, and we have some idea of how it's caused and what differences it makes, but argue about whether particular types of I.Q. test provide good ways of quantifying it. Perhaps, for example, these tests might be challenged as subject to some cultural bias, so that results are affected by the specifics of one's general knowledge in a way that I.Q. itself is thought to be indifferent to. Here, let's suppose, the existence of the thing, intelligence, may not be in question, but particular ways of detecting or measuring it are up for assessment as more or less valid ways of measuring that thing.

The other, more radical dimension of the idea of the validity of a construct has to do with whether there is anything there at all to detect or measure. So, for example, the category 'neurasthenia' might be declared to be 'invalid' not because any one measure of it is somehow incorrect, but because there is no such thing at all. Similarly, in physical chemistry the category 'phlogiston' could be declared to be invalid, not because any measure of it is incorrect, but because it doesn't exist. This can happen even when we're not working with a fully explicit characterization of the variables in the relevant causal structure, but are, for example, thinking of 'phlogiston' as merely a latent variable playing a specified causal role. There may be no latent variable with that causal role.

These two ways of thinking about validity are in practice often twined together, because if you consider DSM categories, there is in practice often no distinction made between the concept and the ways we use to detect and measure the characteristic. If you ask for the meaning of 'schizophrenia', you will be sent to the diagnostic criteria. That can make it seem puzzling how there could be a distinction between the question whether the concept is 'valid' in the sense that those criteria correctly measure the thing, and whether the concept is 'valid' in the sense that it stands for something.

Nonetheless, I think it's important to try to hold on to this distinction, because it gives us a fixed structure we need in addressing the deeply puzzling question of the validity of our diagnostic categories in psychiatry. Let's begin with the question: what more is there to the validity of a concept, in the sense of it standing for something, rather than merely the value of the diagnostic criteria associated with it? Now straight off I think that there are two things we can specify here:

(1) In order for there to be such a thing as schizophrenia, for example, there must be the external phenomenon to which we are causally responding in using the term. There must be something 'out there' we're responding to. That is the condition we're talking about. This is where it seems relevant to talk about a particular type of causal structure, such as, for example, an essentialist structure with a hidden essence causally generating various symptoms.

This condition states a kind of minimal realism that seems to me implicit in talking about the validity of categories in the first place. It's a substantive condition, often denied by

people who say that they take a 'pragmatic' or 'instrumentalist' approach to psychiatry. I'll amplify on this in a moment, but for now I want only to remark that the metaphysical slant these authors put on their discussion usually seems better recast as an epistemic one. These writers are generally moved by a kind of epistemic modesty, a concern that we should not be claiming to know more than we do about the phenomena of mental illness out there. This kind of modesty is better served by realism about disorders coupled with an appreciation of how slender our knowledge of these categories is, rather than by attempting to cut down the reality of mental disorders to fit our current epistemic capacities. The second thing about concepts standing for something is this:

(2) We have to have some conception of what structure it is out there that we're causally responding to. This 'governing conception' that we associate with a term is a picture or model of the kind of phenomenon that we're causally responding to when we use the term.

Richard Boyd explained this idea in a now-famous paper, 'Metaphor and Theory Change' (1979), in which he gave some examples of 'theory-constitutive' metaphors, which can't be eliminated from the practice of science, and which can 'accomplish non-definitional reference-fixing'. Examples he gave from psychology included:

> the suggestion that certain information is "encoded" or "indexed" in "memory store" by "labeling," whereas other information is "stored" in "images";

> disputes about the extent to which developmental "stages" are produced by the maturation of new "programmed" "subroutines," as opposed to the acquisition of learned "heuristic routines," or the development of greater "memory storage capacities" or better "information retrieval procedures";
>
> the view that learning is an adaptive response of a "self-organizing machine";
>
> (Boyd (1979). 360)

Boyd's point was that these kinds of ideas play a role in determining what's being referred to when a scientist talks about 'iconic memory', for example. The conception being used to fix reference here needn't be detailed, and insofar as it is detailed it needn't be more than roughly accurate. It's a matter of having a working model, a picture, a metaphor that can be developed in many different ways. Particularly, it will offer some insight into the causal structure of the phenomenon, by suggesting that the causal structure of one domain can be modeled on the causal structure of another.

Now Boyd confined his remarks to 'relatively mature sciences' (482). You might say that psychiatry doesn't fall under this head. But it seems to me that psychiatrists are typically in practice working with quite rich 'governing conceptions' of the disorders they name, even if those metaphors don't explicitly make it into DSM. The governing conception associated with 'hysteria', for example, can shift over time. And if it finally emerged that clinicians in classifying people as falling or not falling under this category had been responding to the presence or absence of some specific form of PTSD, then we might say: 'Yes, there is such a thing as hysteria, it turned to be ….'. On the other hand,

if it emerged that what clinicians had been responding to was a purely neurological condition, with no significant psychological etiology, we might be more inclined to say: 'People who formerly were classified as 'hysterical' are now thought to have neurological condition X; properly speaking there's no such thing as hysteria'. (Cf. Ahn and Kim 2008 for a review of ways in which clinicians conceptualize the causal structures of disorders.)

These two aspects of a diagnostic concept, the existence of an external phenomenon to which it's a response, and that thing fitting, more or less, our 'governing conception' of what kind of phenomenon that is, seem required if we're to have a diagnostic concept that refers to anything at all. And it's only once our diagnostic concept can be said to be referring to something that it even makes sense to inquire into the validity of our ways of determining whether the characteristic is present in any given case, or of measuring different features of the condition. We might diagram the situation as in Figure 3.

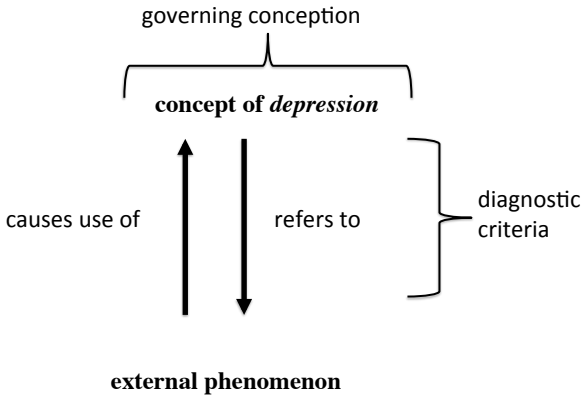

Figure 3. A realist picture of the relation between the concept for a disorder, the external phenomenon it names and the diagnostic criteria for that disorder.

Within philosophy this kind of picture is not remarkable; many authors would endorse something like this structure (cf., e.g., Evans 1982, Fodor 1987, Peacocke 1992, though not all put it in just this way, e.g. Fodor does not explicitly accept a role for anything like 'governing conception'). For our purposes the key point is that a discussion of the validity of a category should be set up in two stages: there is first the question of reference, 'Is there any such thing as X?', does the concept refer at all? If the concept is valid in the sense that there is something it refers to, then we can move to the second stage. Here we can look at particular ways of finding whether something has the characteristic in question, and ask whether they are good ways of doing that.

One thing that is mildly confusing about using this picture in practice is that talk about the 'diagnostic criteria' comes in not only at the second stage, when we are asking whether the criteria are any good. Diagnostic criteria will also matter at the first stage, when we are asking whether there is an external phenomenon out there causing our use of the concept, and to which the concept therefore refers. For the 'causation' of our use of the concept by the external phenomenon will typically be a matter of it triggering our use of the concept by our applying the diagnostic criteria. That does not eliminate the distinction between the two levels. After all, a condition might trigger our use of the concept of 'neurasthenia' when we apply standard diagnostic criteria for neurasthenia; we could therefore say, 'this is the condition we were talking about when we used that category'. It's still open that we might give favorable or unfavorable assessments of how good or bad those criteria are as ways of finding out about that condition. The role

of diagnostic criteria in fixing the reference of a concept doesn't mean that those criteria must automatically be assessed as good ways of finding whether the concept applies.

Let me relate this way of thinking about validity to the way the idea is usually explained to students in science. The concept of validity is often informally explained in terms of the functioning of a gun-sight. A well-functioning gun-sight allows you to identify the target you want to hit, and you can use it to get a bullet to the thing. Analogously, an ordinary concept has associated some idea of what it is you're referring to when you use it, and ways of determining when the concept applies.

In the case of a gun-sight, there are different ways in which the thing might go wrong. Suppose there are a number of possible targets in front of you, and after firing, you can verify how well you've done. Suppose you select a single target using the gun-sight, and take several shots at that thing. There are two questions you can raise:

(a) Do all your bullets land in the same place as one another (whether or not that's the spot occupied by the target)?

(b) On the whole, do your bullets tend to cluster round the intended target, or is the average landing spot for them biased in some direction away from the target?

The first question has to do with whether there is any statistical uncertainty in where use of the gun-sight will get bullets to land; the second has to do with whether there's any systematic uncertainty in where the bullets will land. Or as it's sometimes put, the first

question has to do with the precision of the gun-sight, the second has to do with its accuracy. Or, finally, the first is a question about the reliability of the gun-sight, the second is a question about its validity.

How does this picture apply to, for example, our use of a concept like depression? Well, there is the characteristic out there – the target - that we're responding to, and there are the different ways we have of determining whether the phenomenon is present in a particular case. The whole approach depends on supposing that there is some target that is the thing we're aiming at. Once that is in place, we can ask whether the criteria we're using are good ways of getting onto it, or if one set of criteria is better than another. In the case of a concept like depression, I've in effect been saying that what determines the target is this: which real-world property is it that usually triggers our applying the concept, when we use our current diagnostic criteria? But it's not that we have only the blank fact of our causally responding to that phenomenon: we also have the 'governing conception' that informs our understanding of what we're talking about, and constrains what the concept can be said to be referring to. Getting the right external property in place is the analog of fixing a target; once that is done, we can assess ways of getting onto it as more or less valid. So we can add to our previous picture a step of validating the criteria we use, as ways of getting onto the intended external phenomenon. This is shown in Figure 4.

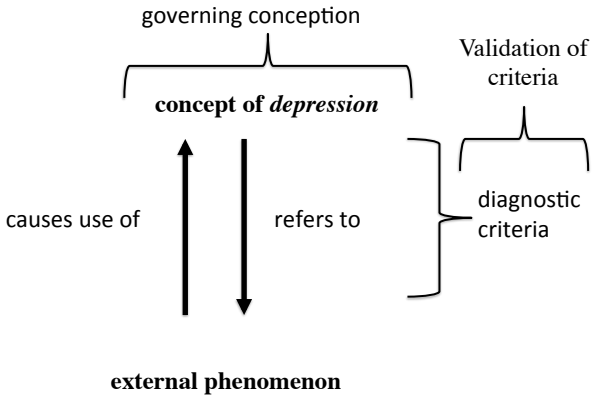

Figure 4. The place of validation of criteria, on a realist picture. The aim of validation is to determine (a) whether the concept identifies any external phenomenon at all, and (b) if so, whether the diagnostic criteria provide a good way of detecting that thing.

So far, this is just like the gun-sight example. The great difference is that in the most basic cases in science, it sometimes seems that we do not have any analogue of looking at the scene directly, without the benefit of the gun-sight. We typically do not have any way, external to the use of the gun-sight itself, of identifying the target or determining how closely our shots have landed. If you think about what the implications of this would be for the assessment of an ordinary gun-sight, they seem radical indeed. On the face of it, we wouldn't have any way of assessing reliability or validity. This is the situation quite dramatically in psychiatry, where there isn't any way, independent of the DSM criteria themselves (or rival criteria that aren't, currently, in any significantly

better an epistemic position) of assessing when we really have found someone with, say, schizophrenia. So let's look at how in practice validation of a DSM category goes.

3. Validation in DSM

From the perspective of those remarks, the central problem with the approach to validation used in DSM is that for the purposes of the exercise of validation, it's assumed that our concepts of each of the disorders are exhausted by the diagnostic criteria. DSM doesn't appeal to the idea of the phenomenon to which we're causally responding, or to our 'governing conception' of the disorder. This is problematic, because it isn't enough to fix the references of the names of the disorders. And if we haven't even attempted to fix the references of the terms, we can't make sense of validity as a distinction between terms that refer and terms that don't, or good ways and not so good ways of spotting the presence or absence of the referent of the term.

To anyone familiar with the way of thinking of validators used in DSM, at least since Robins and Guze (1970), my remarks so far may seem alien to the point of unintelligibility. The guidelines for changes to DSM proposed for DSM V included the following list of types of validators, with the most important asterisked:

I Antecedent Validators
A. *Familial aggregation and/or co-aggregation (i.e., family, twin or adoption studies)
B. Socio-Demographic and Cultural Factors
C. Environmental Risk Factors
D. Prior Psychiatric History
II Concurrent Validators
A. Cognitive, emotional, temperament, and personality correlates (unrelated to the diagnostic criteria).
B. Biological Markers, e.g., molecular genetics, neural substrates
C. Patterns of Comorbidity
[Note - while categories A and B would most typically be assessed after illness onset, they also could be assessed prior to illness onset as pre-morbid characteristics]
III Predictive Validators
A. *Diagnostic Stability
B. *Course of Illness
C. *Response to Treatment

(Kendler et. al. 2009, 27)

Now one thing that is striking about this list is that, on the surface at any rate, it proceeds entirely in terms of correlations. We are looking what correlates we can find with someone's having been established to have a disorder, by means of a particular set of diagnostic criteria. The validity of the disorder has to do with the number and strength of those correlations. And as remarked, there may be different weights given to different types of validator in this process.

This is an intuitive approach, it would hardly have been so widely accepted if it weren't. And I am not going to suggest that there's anything fundamentally wrong-headed about it. But my first point is that it masks the distinction I made above, between

(a) establishing that there is something to which the concept refers (that there is any target there at all), and (b) determining how good or bad our ways are of detecting the presence of that thing, or measuring it.

My second point is that this approach does not explicitly give any place to an analysis of the causal structure of the disorder in determining the validity of a diagnosis. And yet, intelligent use of these criteria of validation seems to depend on some interpretation of causal structure. There are different ways in which we can think of the use of DSM criteria. One is roughly analogous to the way we think of criteria for giving someone a job, such as 'exam qualifications', 'commitment', 'relevant social skills', and so on. Here we need not be thinking that there is some one underlying condition to which all these indicators point. But it would be perfectly possible to assess the validity of a particular picture of a 'strong applicant' by looking at the antecedent, concurrent and predictive validators of the tests that are actually being applied by interviewers; indeed, something like that is how these tests actually are evaluated. On the other hand, you could be thinking of 'the disorder' as something that is the causal outcome of the antecedent validators, which is expressed in the concurrent validators, and has as a causal outcome the way things go with the predictive validators. This is quite different to the case of criteria for a successful job application. On the face of it, it seems entirely possible that one conception of the causal structure here may lead to a quite different weighting of validators than does the other conception of causal structure. It's hard to see how the discussion of validation in DSM can really operate at the level of correlations, without at least implicitly bringing in some picture of the causal structure of the disorder.

So I've made two comments about the Robins and Guze-style criteria. One is about the need to separate out the 'reference-fixing' dimension of validity – does the term refer to anything at all? – from the other dimension: given that the term refers to something, how could are our diagnostic criteria for detecting the presence of that thing? The other comment was about the need to operate with some picture of the causal structure of the disorder – are the criteria for the disorder actually constitutive of possession of the disorder, for example, or are they intended to point to some further inner condition? But these two comments are, I think, not independent. The 'governing conception' that we have of a disorder will in the first instance be an interpretation of the causal structure of the disorder. Until we have determined what kind of causal structure we're looking for, we haven't identified a condition at all. To put the point in a more confrontational way, the current combination of a DSM that merely lists diagnostic criteria, coupled with a Robins and Guze-style approach to validation, can't actually succeed in fixing references for the names of disorders at all. If that's all we have to go on as fixing the reference of a term like 'major depression', for example, then what does it take to be depressed? Is it enough if one merely meets the diagnostic criteria? Or is there some further condition that one has to have, so that one could in principle meet the diagnostic criteria and yet not in fact be depressed (just as one could have the symptoms of a viral illness yet not have the virus)? If you try to keep the chips up on that kind of question, claiming that our knowledge is as yet insufficient, then you haven't fixed a reference for the term 'major depression' at all. You have missed out a crucial component in the attempt to fix a reference for your term. Consequently, there's no saying whether there is or is not such a thing as 'depression'. And there's little point in

trying to raise the question whether your current tests are good indicators of the presence of that thing. What thing?

In fact, the evidence seems to be that clinicians do not interpret disorders in ways that stay out of the question as to their causal structures. Kim and Ahn (2008) found that clinicians generally had clear causal structures associated with the diagnostic criteria for disorders, and hypothesized that 'symptoms that cause many other symptoms (i.e., causally central) would be treated as being more important than symptoms that cause few other symptoms (i.e., causally peripheral)' (2008, 4-5). They remark,

> In contrast [in DSM], the system is set up, with a few explicit exceptions, so that all symptoms in a given disorder are equally weighted. For instance, the four symptoms with boldface boxes in Figure 2 [Figure 5 below] must *all* be present to warrant a diagnosis of Anorexia nervosa, making all four symptoms equally important for classification. However, according to the clinicians' data collected in our experiments, "distorted body image" was most causally central in the clinicians' theories, whereas "absence of the period (in women) for more than 3 menstrual cycles" was rated the most causally peripheral. Furthermore, "distorted body image" was considered to be the most diagnostically important of the criteria, and "absence of the period (in women) for more than 3 menstrual cycles," though also a *DSM* diagnostic criterion for Anorexia nervosa, was considered to be the least diagnostically important. We obtained similar patterns of results across eight other mental disorders (Kim & Ahn, 2002).
>
> (2008, 5)

They suggest that "instead of sticking with a purely descriptive approach, incorporating a causalist approach, whenever possible and wherever reasonable, may actually encourage clinicians to rely more on the *DSM*. As we suggested, incorporating causal information at the symptom-to-symptom level might be a reasonable place to start. Attempting to adhere solely to a descriptive approach in the DSM may not necessarily lead to better reliability in clinicians' diagnoses." (2008, 7).

These points, if well-taken, imply that it is not as if there is some pragmatic value in keeping causal information out of DSM. And what I have been saying previously is that unless causal information is allowed into what I've been calling our 'governing conception' for a disorder, we simply can't claim to have uniquely specified any real-world phenomenon. I think it is this failure of the DSM names, if strictly and literally interpreted merely in terms of their associated criteria, to actually identify any particular condition that makes the whole question of their validation so confusing. It is also what encourages the frequent claims by people reflecting on DSM that the diagnostic categories are merely 'social constructs'. If the terms don't refer to anything 'out there', then perhaps they do only have to do with some projections of the theorists. But that's a mistake. The problem is that the terms are being interpreted so that they don't refer at all, rather than that they refer to 'social constructs', whatever they are.

Figure 2

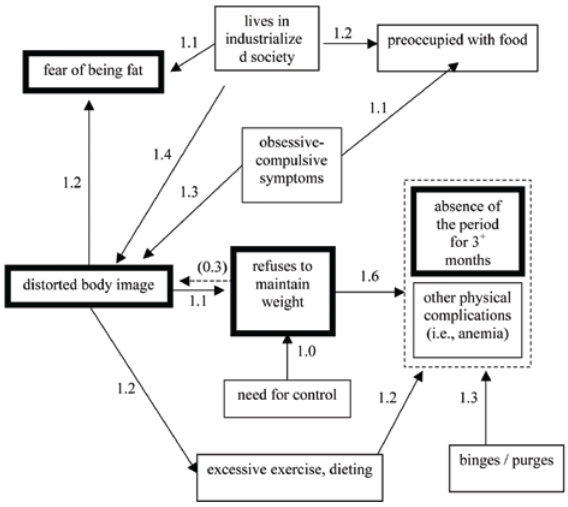

Figure 5. Composite of clinicians' drawing of Anorexia Nervosa. From Ahn and Kim 2008. Notice that this is not an essentialist picture, nor a 'homeostatic cluster' picture. But it does inform the understanding and use of the diagnostic criteria.

I think that the official picture of validation in DSM can be seen as a kind of amputated form of the picture of validation that I gave in section 2 above. Let me be explicit that I think the amputation was well-motivated: the motive was great, and well-grounded, epistemic caution. But I think it has now been completely forgotten that this was an amputation of the usual picture of validation. We can diagram the DSM picture of validation as in Figure 6 below.

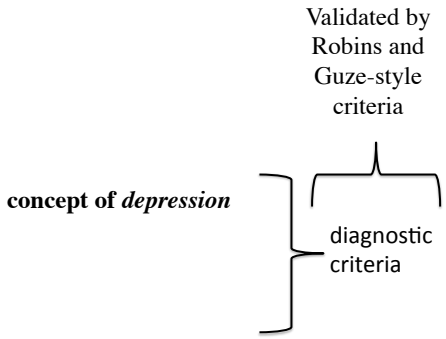

Figure 6. A picture of validation that tries to do without the idea of the 'external phenomenon' to which the concept refers, so that we don't think of validation as involving either (a) determining that there is such a thing, or (b) whether the diagnostic criteria are good ways of detecting it.

Compare this to Figure 4 above. From this picture, anything about our 'governing conception' of the causal structure of the disorder had been dropped. And the very idea of an external phenomenon out there, to which we are causally responding and to which we are referring, is doing no work in the picture. We saw that a common explanation of the usual scientific picture of validation is as a matter of determining how good the gun-sight is, whether we're getting onto a target with it at all, and if so, how good the gun-sight is at doing that. The official DSM picture is like that, only the very idea of the target has been dropped, and the idea of the shooter having sight of the target, so that they

know which thing they're trying to hit, has been abandoned. All we have are correlations with signs that the gun is about to be used, concurrent and predictive corrrelations with the use of the sight. If these all seem to pan out, then we say that the gun-sight has been validated.

Now the point about the amputated conception of validity is that it ultimately makes sense only when you realize that it is an amputation. The whole point of the validating criteria is to try to find when you are getting onto the external phenomenon. You might be cautious about what you claim to know about the validation of the gun-sight, but the whole point of the exercise is in the end to find out whether you are hitting anything. If you don't keep in mind that this is ultimately what's going on, then it is likely to seem to you that the whole business of gun-sight validation is nothing more than a 'social construct' or something like, because what on earth is this ritual of validation about, if it isn't about finding out whether you're hitting something?

Psychiatry has been using Robins and Guze-style validators for decades, and a vast body of important knowledge has been developed using them. You can live within this world. But that means that when you ask: do I need the notion of 'the disorder itself', the external phenomenon to which we're causally responding and to which I'm referring?, well that notion does not seem to be part of the official validation process. So at this point you might say: do I need the notion at all? Perhaps all it comes to that 'there is such a thing' as schizophrenia, for example, is that the validation exercise has passed our diagnostic criteria as valid. There isn't any further question. Our concept of schizophrenia, on this way of thinking about it, is exhausted by the diagnostic checklist, and its validity is simply a matter of the Robins and Guze criteria being met. There is no

more to the existence and nature of schizophrenia than that. It's only a 'social construct', a product of our diagnostic and validatory practices. There is something quite paradoxical about this position. Remember that it began with great epistemic caution, a sense that we do not know much about the disorders. But now the amputation of our validatory practices leads us to a picture on which there isn't any more to know about schizophrenia, for example, than what we currently know. After all, if the thing is merely a social construct, a product of our diagnostic and validatory practices, how could there be any more to know about it than those practices themselves generate? We have lost the very picture of the disorder 'out there', the thing about which we know little, that is what motivated the original amputation of our process of validation.

It should be evident that this is all a by-product of losing sight of the bigger picture that can alone make sense of validation. We need the idea of the thing out there to which we're referring, if we're to make sense of what we're doing in validation. That means that we have to reckon in to our concept of schizophrenia not just the diagnostic criteria, but our 'governing conception' of the thing, our grasp of its causal structure, and the disorder out there to which we are causally responding in giving diagnoses. As we've seen, that would already reflect good clinical practice. The trouble is that when we try to think how we would go beyond the approach of Kim and Ahn to find true causal models of disorders, we run into the 'Princess Elisabeth' problem. We do not know what kinds of mechanism and process we should be thinking in terms of when we reflect on the causal structures of disorders.

ACKNOWLEDGEMENT

Such understanding of these issues as I have, I owe almost entirely to Ken Kendler. I was also helped by discussion of an earlier draft at the Copenhagen meeting and by Dominic Murphy's comments there. In my understanding of how reliability and validity are thought of in the physical and social sciences, I have been helped by discussions with Robert MacCoun and Saul Perlmutter.

REFERENCES

Ahn, Woo-Kyoung and Nancy S. Kim. 2008. 'Causal Theories of Mental Disorder Concepts'. *Psychological Science Agend*a, 22, 3-8.

Boyd, Richard. 1979. 'Metaphor and Theory Change'. In Alex Ortony (ed.), *Metaphor and Thought*. Cambridge: Cambridge University Press.

Dowe, Phil. 2000. *Physical Causation*, Cambridge: Cambridge University Press.

Evans, Gareth. 1982. *The Varieties of Reference*. Oxford: Oxford University Press.

Fair, David. 1979. 'Causation and the Flow of Energy', *Erkenntnis* 14, 219-250.

Fodor, Jerry. 1987. *Psychosemantics*. Cambridge, Mass.: MIT Press.

Hoerl, Christoph. 2013. 'Jaspers on Explaining and Understanding in Psychiatry'. In Thomas Fuchs and Giovanni Stanghellini (eds.), *One Century of Karl Jaspers' General Psychopathology*. Oxford: Oxford University Press.

Hume, David. 1740/1975. *A Treatise of Human Nature*, edited by L. A. Selby-Bigge, 2nd ed. revised by P. H. Nidditch. Oxford: Oxford University Press.

Jaspers, Karl. 1923/1963. *General Psychopathology*. Chicago: University of Chicago Press.

Kendler Kenneth S., D. Kupfer, W. Narrow, K. Phillips and J. Fawcett. 2009. 'Guidelines for Making Changes to DSM-V'. Washington, DC: American Psychiatric Association. Unpublished manuscript.

Locke, John. 1690/1975. *An Essay Concerning Human Understanding*, ed. P.H. Nidditch. Oxford: Oxford University Press.

Mattern, Ruth. 1978. 'Descartes's Correspondence with Elisabeth: Concerning Both the Union and Distinction of Mind and Body'. In Michael Hooker (ed.), *Descartes: Critical and Interpretive Essays*. Baltimore: Johns Hopkins University Press.

Robins, E. and S.B. Guze. 1970. 'Establishment of Diagnostic Validity in Psychiatric Illness: Its Application to Schizophrenia'. *American Journal of Psychiatry*, 126, 983-987

Snow John. 1965. *Snow on Cholera: Being a Reprint of Two Papers by John Snow With a Biographical Memoir*. New York: Hafner.

Peacocke, Christopher. 1992. *A Study of Concepts*. Cambridge, Mass.: MIT Press.