Interventionist causal models in psychiatry: repositioning the mind–body problem

K. S. Kendler^{1,2*} and J. Campbell³

¹ Virginia Institute of Psychiatric and Behavioral Genetics, and ² Departments of Psychiatry, and Human and Molecular Genetics, Medical College of Virginia/Virginia Commonwealth University, Richmond, VA, USA ³ Department of Philosophy, University of California, Berkeley, CA, USA

The diversity of research methods applied to psychiatric disorders results in a confusing plethora of causal claims. To help make sense of these claims, the interventionist model (IM) of causality has several attractive features. First, it connects causation with the practical interests of psychiatry, defining causation in terms of 'what would happen under interventions', a question of key interest to those of us whose interest is ultimately in intervening to prevent and treat illness. Second, it distinguishes between predictive-correlative and true causal relationships, an essential issue cutting across many areas in psychiatric research. Third, the IM is non-reductive and agnostic to issues of mind–body problem. Fourth, the IM model cleanly separates issues of causation from questions about the underlying mechanism. Clarifying causal influences can usefully structure the search for underlying mechanisms. Fifth, it provides a sorely needed conceptual rigor to multi-level modeling, thereby avoiding a return to uncritical holistic approaches that 'everything is relevant' to psychiatric illness. Sixth, the IM provides a clear way to judge both the generality and depth of explanations. In conclusion, the IM can provide a single, clear empirical framework for the evaluation of all causal claims of relevance to psychiatry and presents psychiatry with a method of avoiding the sterile metaphysical arguments about mind and brain which have preoccupied our field but yielded little of practical benefit.

Received 19 February 2008; Revised 5 August 2008; Accepted 5 September 2008; First published online 10 October 2008

Key words: Causality, interventionist model, mind-body, psychiatry.

Introduction

Research in psychiatry generates a bewildering diversity of causal claims. A reader of any of our major journals such as this one confronts efforts to prove that the risk for psychiatric disorders is influenced by economic, cultural, psychological, genetic, neural, brain structural and molecular influences. Underlying this broad range of casual claims is the fault-line of the mind-body problem. How do causes of psychiatric illness best understood as mental processes interrelate with those that reflect physical or biological events? Common sense suggests that causal pathways exist in both directions between the mind and the brain. However, despite much effort, the nature of such casual paths has remained frustratingly obscure.

Insufficient attention has previously been paid to the nature of causal claims in psychiatry. We argue that recent advances in the philosophy of science – particularly in the nature of causation and explanation – are of substantial relevance to our field. Specifically, interventionist causal models can provide a framework for the field of psychiatry that applies equally to the broad array of potential causal processes at work in psychiatric illness, including those on either side of the mind-brain 'divide'. Interventionist models (IM) show how causal variables of different types can figure in a single rigorous account of the causation of psychiatric illness.

Experiment is usually taken to be the acid test for a causal hypothesis. The IM reflects this understanding. It provides an explicit characterization of the type of experiment that is ideally required to determine the correctness of a causal claim. As such, it may be helpful to have this framework in mind when interpreting specific results.

The interventionist approach explains what it is for a causal connection to hold in terms of the results of an idealized experiment. There is in principle no particular limitation on what combinations of variables might be used on this approach to characterize the causes of a disorder. So the approach makes explicit how we could have genuinely 'multi-level' explanations in psychiatry, the kind of models increasingly favored among philosophers of science for

^{*} Address for correspondence : K. Kendler, M.D., Virginia Institute for Psychiatric and Behavioral Genetics, Virginia Commonwealth University Medical School, Box 980126, 800 E. Leigh Street, Room 1-123, Richmond, VA 23298-0126, USA.

⁽Email: kendler@vcu.edu)

complex natural phenomenon like psychiatric illness (Schaffner, 1994; Gold & Stoljar, 1999; Craver, 2007). Interventionism allows variables of many different types – psychological, economic, and biological, for example – to be used together in stating the causes of a disorder. The IM is also a needed antidote to the inclusiveness of these models which can be misinterpreted as an uncritical holistic approach in which everything is deemed relevant. While numerous causal processes are likely to be of etiologic importance for most psychiatric disorders, their inclusion in multi-level causal processes cannot be assumed *a priori*. Rather, each must be independently vetted to show that it possesses real causal information. The IM is an ideal framework within which to evaluate these claims.

The IM can also help formulate the issues around the mind-problem in a novel way, one that may be more fruitful for psychiatry than traditional metaphysical approaches. The idea is this. Interventionism says that we can use experiment to demonstrate the existence of a causal connection between a causal factor C and an outcome or effect E. When we have such a demonstration to hand, there is a further connected but discrete issue: what is the mechanism by which C and E are related? Prior approaches toward the mindbody problem have tended to confound these two levels - that of demonstrating causal influence and that of clarifying the mechanism that instantiates that causal influence in terms of the specific biological and/or psychological processes. The IM cleanly separates these two questions. The elucidation of causal relationships in the mind/brain is thereby freed from the heavy metaphysical baggage (and the deep scientific questions) that comes with trying to specify the particular mechanisms that might underlie the causal relationships in the mind/brain system.

Finally, the IM is more demanding than scientific frameworks that seek merely for successful prediction. While a barometer reading will predict the weather, changing pressure readings in a barometer by artificial means will not alter weather patterns. The IM captures, in a stark manner, the difference between correlation and causation, an issue central to nearly all of psychiatric research.

In what follows, we first set out the IM. Then we look at its relation to an approach that thinks of causation in terms of 'mechanisms'. Finally, we look at further problems and prospects for the IM of causation.

The Interventionist Model

How does the IM work? Consider a simple, idealized case. Suppose we want to determine whether stress (S) increases the risk for major depression (MD). An 'ideal

experiment' here would be the unethical one in which, in a given population, we randomly intervene on individuals exposing them to a stressful experience such as public humiliation (H). This experience increases their level of S and we observe if they subsequently suffer from an increased incidence of MD. Our design is:

H intervenes on $S \rightarrow MD$.

Thus, we are assuming that intervention on stress will make a difference to risk for MD. For this to work, though, the 'intervention' must meet a number of conditions. An explicit formulation has been proposed (Woodward & Hitchcock, 2003; Pearl, 2000) which we illustrate with our thought experiment. In what follows we use our thought experiment to illustrate the analysis proposed by Woodward & Hitchcock (Pearl, 2000).

- (1) In the class of individuals who are and are not exposed to our intervention, H must be the only systematic cause of S (so that all of the averaged differences in level of S in each cohort of our exposed and non-exposed subjects result entirely from the humiliation).
- (2) H must not affect the risk for MD by any route that does not go through S (for example by causing individuals to stop taking their antidepressant medication).
- (3) H is not itself influenced by any cause that affects MD via a route that does not go through S as might occur if individuals prone to depression were more likely to be selected for humiliation.

The IM says that questions about whether X causes Y are questions about what would happen to Y if there were an intervention on X. National threats, such as the events of 9/11, can be considered to cause reduced rates of suicide in the threatened nation (Salib, 2003) because, had there been an intervention to prevent those threats, the suicide rate would not have changed.

Let us apply the IM to another common empirical question in psychiatry: Do the levels of a particular drug (D) in an individual reduce the symptoms (Sxs) of schizophrenia? We need to develop an intervention that can isolate the effect of our putative C – levels of D – on our E – here Sxs of schizophrenia. One obvious way to do this is to randomize (R) exposure to D in a population. Our design would look like

R intervenes on $D \rightarrow Sxs$.

What we have in fact produced is a randomized controlled trial which can be understood as the practical realization of an IM of causal inference. Note that the model dictates care to insure that the only systematic effect on drug levels in our study results from the randomization. If subjects have access to the drug on their own, that could confound the results. Also note that the causal effect of R on Sxs must flow through D. This makes explicit the need for placebo control. If taking the active drug D impacts on Sxs through a variable other than D, we lose our ability to infer the causal effect of D on Sxs.

The point we want to emphasize about the interventionist approach is that it allows psychiatrists freedom to use whatever family of variables seems appropriate to the characterization of a disorder. There is no assumption that the variables have to be capable of figuring in general laws of nature or that the variables have to relate to some specific physical mechanism. The fact is that the current evidence points to causal roles for variables of many different types, and the interventionist approach allows us to make explicit just what those roles are.

For all that, the approach is completely rigorous. Though our exposition here is highly informal, we are providing an intuitive introduction to ideas whose formal development has been vigorously pursued by others (especially Spirtes *et al.* 1993; Pearl, 2000 and Woodward, 2003).

Mechanisms

Suppose we find a correlation between two factors X and Y, say smoking and cancer, and we wonder whether X is a cause of Y. After all, there are other possibilities: there might be some common cause of X and Y, for example. What is required for there to be a causal connection, rather than a mere correlation, between X and Y? The IM says: what is required is that interventions on X should make a difference to the value of Y. There is, however, another natural thought that arises. What is required for the causal connection between X and Y? That is, what is the specific mechanism connecting X and Y? Our main point in this section is that the idea of a mechanism should not be thought of as providing a competitor to the IM, but rather a supplementation to it. Moreover, some of the most puzzling methodological issues in psychiatry can be stated in terms of the problem: what 'notion' of a mechanism should we be using?

The simplest notion of 'mechanism' is that of a specific and definable spatio-temporal or physical process connecting cause and effect. This model is exemplified by the contact between a moving and a stationary billiard ball. Momentum is transferred from the first to the second ball through direct physical contact as signified by the satisfying 'click'. Notice that the 'mechanical model' says more than merely that we are looking for 'causal mechanisms'; it insists that causality is fundamentally a matter of spatio-temporal

contact. Could we use this idea of 'mechanism' to explain what causation is, without appealing to the notion of intervention at all?

We think not and will illustrate this key point with several examples. Consider the empirical observation that the rate of suicide declined significantly in England in the weeks after September 11, 2001 (Salib, 2003). To explain this phenomenon from a mechanistic perspective on causation requires that we delineate a spatio-temporal process that connected the events of 9/11 in the United States to people in England. For example, information about the events of 9/11 was conveyed to the English population through a variety of mechanisms including radio, television, e-mail, word of mouth and newspapers. From an interventionist perspective, though, it does not follow from this point that the causal variables that we use to describe what happened must themselves be explicitly defined in terms of those physical processes. Surely, focusing on the particular medium of information transfer is the wrong level at which to understand the changes in suicide rates. Rather, credible causal accounts will have to invoke constructs such as 'external threats', 'increased social cohesion', and 'reduced feelings of alienation ' that cannot be easily understood at a mechanical or physical level (Durkheim, 1897).

Consider a simple physical example - the expansion of a balloon as the gas inside it is heated. The mechanical model of causation would explain this process as the sum of the action of individual gas molecules, accelerated through heating, striking the balloon. While each molecule contributes to the expansion of the balloon, this is not the best level at which to understand the process. The information contained in the trajectory of any individual molecule is not what matters for whether or not the balloon expands. The balloon expansion is better understood as a result of a higher level process – the average kinetic energy of all the gas molecules caused by their heating. To intervene in the process - reduce or accelerate the expansion-would mean to turn up or down the heat - not attempt to alter the trajectory of individual molecules.

Consider finally a monkey learning a perceptual discrimination. When the monkey hears a high-*versus* a low-pitched tone, if it presses the red square instead of the green triangle on the computer screen, it gets 5 ml of apple juice. The tone certainly 'causes' the motor response, in that whether the response occurs depends on whether the tone does. There are of course many intervening factors in the process. But these intervening factors on which the occurrence or non-occurrence of the motor response depends have to be understood at many different levels. The nature of this causal cascade cannot be efficiently or even

realistically understood as a basic mechanical process. Information, encoded in action potentials of large groups of neurons, passes through a complex series of neural pathways involving pitch perception of the auditory stimulus, shape, shading and color of the visual stimulus, memory of the rewarding properties of the juice and coordination of the motor outflow pathways to the monkey's arm. The causal pathway reflects a series of interacting neural pathways each composed of many neurons having thousands of receptors and ion channels through which in turn billions of cations flow back and forth. Whether or not a particular motor response occurs does not depend on the occurrence of any particular mechanical event in this process, such as a sodium ion passing through a membrane. Causation in complex biological systems that contain interacting systems spanning several levels of complexity cannot typically be easily understood as a simple, mechanical process. Although simple mechanical processes are clearly involved, important causal effects are typically seen at higher levels such as neural systems (see Briggman et al. 2005 for a particularly clear example).

Seeking to understand causal mechanical processes in psychiatric illness at a basic biological level is a highly valuable research approach that has yielded important insights. This method cannot, however, be extrapolated to many important classes of causal variables in psychiatry and so is too restrictive to be used as a general approach to causation in our field. Critically, the IM can work independently of any underlying assumptions about the specific causal mechanisms involved. Results from the IM can isolate the problem - the existence of relationships between a particular cause and a specific effect - and thereby guide the empirical research. But the conclusions reached in the IM are in no way dependent on the degree to which particular biological or psychological mechanisms are at work. The value of the interventionist standpoint is not that it immediately resolves these complex issues of casual mechanisms. It does not. But it allows us to be explicit about their location. Clarifying these deep puzzles of mechanisms is not required to answer the question of causal influence which has to do only with which variables are such that manipulating them makes a difference to an outcome.

Types of intervention

Let us return to the problem of the relationship between stress and MD. While we can in principle actually randomize exposure to medication, we cannot regard the ordinary stressors encountered in everyday life as constituting interventions on stress, in the sense



Fig. 1. Two possible interrelationships between humiliation (H), stress (S), major depression (MD), and personality (P). Arrows reflect putative casual influences. In panel (*a*) the relationship between H, S and MD is causal while in panel (*b*) it is non-causal.

explained above. If we were just to measure the association between stress and MD in a representative population, we would be at risk of violating rule 3 articulated above. To be specific, as depicted in Fig. 1*a*, the correlation observed between humiliation, stress and MD could be entirely causal with personality (P) playing no role in the key relationships. Alternatively, as depicted in Fig. 1b, the correlation observed between humiliation, stress and MD could be entirely non-causal, both resulting from a particular personality (P) which both increases the chances of exposure to humiliating stress and augments the risk for MD. [Indeed high levels of the personality trait of neuroticism increase exposure to stressful events (Kendler et al. 2003) and elevate the risk for MD (Kendler et al. 1993).]

So, we cannot be confident that in the general population, S is causally related to MD. But we are not helpless to approximate a true intervention. Monozygotic twins who are reared together have very similar personalities (Loehlin, 1992). By studying the relationship between H and MD in such pairs, we can come close to 'breaking' the confounding effect of P on the relationship of H and MD. Used in this way, this co-twin control design can be understood as a 'natural intervention'. Within each cohort (here a pair of twins), we have, to a first approximation, the same P. So, when there is a difference in S between the two members of a pair, we have some basis in assuming that this results from an outside influence such as a 'fateful' or 'independent' stressful life event and not from P. So that outside factor constitutes our 'natural intervention' on S. Then we are looking to see whether our S predicts MD under those natural interventions. (See Kendler *et al.* 1999, for an implementation of exactly this design in an attempt to clarify the relationship between S and MD.) The IM permits us to appreciate the power of a co-twin control study which, while not providing us with the confidence of a true intervention, nonetheless can substantially increase our confidence in our causal attributions.

As recently reviewed by Rutter (2007), a number of other natural experiments are available to help us infer causality in biomedicine. These approaches are typically easily understood within an IM framework. One particularly attractive method is Mendelian randomization (Ebrahim & Davey, 2008). In a psychiatric application of this approach, Irons and colleagues evaluated the hypothesis that alcohol use (AU) is a gateway to drug use (DU) in Asian adolescents (Irons *et al.* 2007). Genotype (G) at the aldehyde dehydrogenase 2 locus was the intervention, which would look like this:

G intervenes on $AU \rightarrow DU$.

Thirty percent of the sample, for genetic reasons, produced a deficient enzyme which resulted in unpleasant side-effects after drinking. These subjects consumed much less alcohol (confirming that G intervened on AU) but did not have reduced levels of drug use. These findings disproved the gateway model as the intervention of AU on DU was not effective.

Explanatory processes

Compare the generalizations: 'Divorce increases the risk for MD', and 'Stressful life events cause the risk for MD'. One reason you might prefer the latter is that it seems more general. But this is not the only reason it is preferable; it also suggests that divorce causes MD by causing stress. Imagine that we develop a measure of stress that is shared in common by a wide array of depressogenic experiences including stressful life events, but also chronic difficulties, minor hassles and those idiosyncratic events (loss of a minor object of great emotional significance) that fall 'below the radar' of our research instruments. This will provide a deeper explanation and illustrates the closely related concepts of causal proximity and explanatory depth. Fig. 2 illustrates this in a general way, assuming specific causes C1-C4 which all acted through a common mechanism captured by C_C.

A characterization of the causal process here that identifies the common mechanism C_C is superior to



Fig. 2. The relationship between a range of causes (C_1-C_4) which all acted through a common causal mechanism (C_C) on an effect (E).

any description that merely identifies each of the distal causes C_1 – C_4 . Such a characterization will allow us to understand, for example, why each of the generalizations linking C_1 to E, C_2 to E and so on has certain exceptions. There will be an exception to the generalization if there is some context in which C_i does not cause C_C .

A critical strength of the IM for psychiatry is its nonreductive character. It makes no *a priori* judgment on the level of abstraction on which the causal processes can be best understood. The IM only requires that at whatever level it is conceived, the cause makes a difference in the world. This is so important that it deserves repeating. The IM provides a single, clear empirical framework for the evaluation of all causal claims of relevance to psychiatry. Further, it helps us to separate, into two clear steps, the clarification of casual processes and the search for mechanisms. The clarification of mechanisms is often (but not always) the harder task but lack of understanding of specific mechanisms need not retard our ability to elucidate specific causal claims.

The IM can even go further in providing a framework for addressing a central conceptual issue within psychiatry and the neurosciences more generally – the concentration of causal power at the level of mind *versus* at the level of brain. Let us sketch two possible scenarios for the explanatory pathways from humiliation to MD. In the first - or mental model - the essence of humiliation that is depressogenic can be well measured using first-person subjective reports. However, there may on this model be no unified 'biology of humiliation' that is sufficiently consistent across individuals to give much explanatory power. This is depicted in Fig. 3a where H_M refers to the mental construct of humiliation and H_{B1}-H_{B4} refer to four of the many different brain processes that are consequent upon humiliation in different individuals. In this case, the IM would suggest that the causal interventions be focused at the mental level. In the



Fig. 3. The possible relationship between mental (subscript M) and brain-based (subscript B) models of humiliation and major depression (MD). In panel (*a*) H_M refers to the mental construct of humiliation and H_{B1} – H_{B4} refer to four of the many different brain processes that are consequent upon humiliation in different individuals. In panel (*b*) H_B refers to the single brain effect of various humiliating experiences while H_{M1} – H_{M4} refer to four of the many different mental instantiations of humiliation.

second or brain model, the essence of humiliation is very hard to measure using first-person reports because the nature of humiliating experiences vary so widely from person to person and from culture to culture. We may have no confidence that there is any subjective unity to this wide diversity of mental experiences. Nonetheless, it may be that there is a single, easily assessed brain state that they all cause; and that this is implicated in MD. The causal pathways in this model is depicted in Fig. 3b, where H_B refers to the single brain effect of these various experiences and H_{M1}-H_{M4} refer to four of the many different mental instantiations of humiliation. In this case, the brain state clearly provides the mechanism by which humiliation causes depression and would be the preferred site for intervention. This example demonstrates how within the context of the mind-body problem understanding the impact of specific interventions can frame questions about the specific mechanisms involved.

Limitations of the IM

Two potential limitations of the IM deserve emphasis. We have explained the application of IM to

populations: to general claims such as 'stress causes depression'. In principle there is no reason why the IM should not be applicable to causation in particular cases; to the analysis of claims such as 'John's divorce caused his depression.' The guiding idea will be that for John's divorce to have caused his depression is for it to be that an 'intervention' (in something like the sense we explained) that prevented his divorce would also have prevented his depression. Of course, in particular cases it may be difficult to know whether such an argument is in fact true. But, in principle, the IM is applicable to single cases. For example, if we wanted to know whether drug X improves Andrew's psychotic symptoms, we could repeatedly switch Andrew from drug X to placebo and back again. If his psychotic symptoms consistently improved on drug treatment and then exacerbated when switched to placebo, our confidence in this causal attribution - although involving a single individual - could be quite high.

Second, we have not here discussed the concept of background conditions. For psychiatric phenomenon, the impact of interventions as defined by the IM will surely be conditional on some aspects of the situation. These relevant background conditions may vary widely from genotype (as seen in genotype × environment interaction) to culture (where the same objective life event induces different levels of humiliation because of distinct societal beliefs). These background conditions will introduce a necessary limitation by defining the boundaries of the causal relationship that can be imputed but do not otherwise disturb the central tenets of the interventionist perspective. Where we can, of course, in practice investigators should always try to make these background variables explicit.

Summary

In closing, we emphasize the critical attractive features for psychiatry of the IM. First, the IM is anchored in the practical and reflects the fundamental goals of psychiatry which are to intervene in the world to prevent and cure psychiatric disorders. Second, in asking what would have happened if things had been different, the IM, at its core, distinguishes between correlative and causal relationships - a fundamental issue in much of psychiatric research. Third, the IM provides a single, clear empirical framework for the evaluation of all causal claims in psychiatry. It offers a way by which different theoretical orientations within psychiatry can be judged by a common metric. In so doing, it can provide needed rigor for multi-level causal models which are at risk of an uncritical holism in including all possible risk factors regardless of their level of empirical support. Fourth, the framework provided by the IM can help us find, among the many possible, the optimal level for explanation and ultimately for intervention and assist in structuring our search for causal mechanisms. It can also help us judge the depth and quality of different explanatory systems. Finally, the IM is explicitly agnostic to issues of the mind-body problem. In particular, it permits the clear separation of causal effects from the mechanistic instantiations of those effects. In so doing, the application of the IM can help us replace the sterile metaphysical arguments about mind and brain which have yielded little of practical benefit with productive empirical research followed by rigorous conceptual and statistical analysis.

Acknowledgments

We are, in this essay, indebted to the ideas expounded in Woodward (2003) and their sources in Pearl (2000) and Spirtes *et al.* (1993).

Declaration of Interest

None.

References

- Briggman KL, Abarbanel HD, Kristan Jr. WB (2005). Optical imaging of neuronal populations during decision-making. *Science* 307, 896–901.
- Craver CF (2007). Explaining the Brain. Clarendon Press.
- **Durkheim E** (1897). *Le Suicide 1897* [English translation, 1951]. Glencoe, IL: Free Press.
- Ebrahim S, Davey SG (2008). Mendelian randomization: can genetic epidemiology help redress the failures of observational epidemiology? *Human Genetics* **123**, 15–33.

Gold I, Stoljar D (1999). A neuron doctrine in the philosophy of neuroscience. *Behavioral and Brain Sciences* 22, 809–869.

- Irons DE, McGue M, Iacono WG, Oetting WS (2007). Mendelian randomization: a novel test of the gateway hypothesis and models of gene-environment interplay. *Developmental Psychopathology* **19**, 1181–1195.
- Kendler KS, Gardner CO, Prescott CA (2003). Personality and the experience of environmental adversity. *Psychological Medicine* **33**, 1193–1202.
- Kendler KS, Karkowski LM, Prescott CA (1999). Causal relationship between stressful life events and the onset of major depression. *American Journal of Psychiatry* 156, 837–841.
- Kendler KS, Neale MC, Kessler RC, Heath AC, Eaves LJ (1993). A longitudinal twin study of personality and major depression in women. *Archives of General Psychiatry* 50, 853–862.
- Loehlin JC (1992). Genes and Environment in Personality Development. Sage Publications: Newbury Park, CA.
- **Pearl J** (2000). *Causality Models, Reasoning, and Inference.* Cambridge University Press: Cambridge.
- **Rutter M** (2007). Prodeeding from observed correlation to causal inference: the use of natural experiments. *Perspectives on Psychological Science* **2**, 377–395.
- Salib E (2003). Effect of 11 September 2001 on suicide and homicide in England and Wales. *British Journal of Psychiatry* 183, 207–212.
- Schaffner KF (1994). Psychiatry and Molecular Biology: Reductionistic Approaches to Schizophrenia in Philosophical Perspectives on Psychiatric Diagnostic Classification. Johns Hopkins University Press: Baltimore.
- Spirtes P, Glymour C, Scheines R (1993). Causation, Prediction and Search. Springer-Verlag: New York.
- Woodward J (2003). *Making Things Happen*. Oxford University Press: New York.
- Woodward J, Hitchcock C (2003). Explanatory generalizations, Part I: a counterfactual account. *Nous* **37**, 1–24.