Psychological Medicine

cambridge.org/psm

Top-down causation in psychiatric disorders: a clinical-philosophical inquiry

Kenneth S. Kendler[1,2] and James Woodward[3]

[1]Virginia Institute for Psychiatric and Behavioral Genetics, Virginia Commonwealth University, Richmond, VA, USA; [2]Department of Psychiatry, Virginia Commonwealth University, Richmond VA, USA and [3]Department of History and Philosophy of Science, University of Pittsburgh, Pittsburgh, PA, USA

Review Article

Cite this article: Kendler KS, Woodward J (2021). Top-down causation in psychiatric disorders: a clinical-philosophical inquiry. *Psychological Medicine* **51**, 1783–1788. https://doi.org/10.1017/S0033291721001811

Received: 27 October 2020
Revised: 16 March 2021
Accepted: 21 April 2021
First published online: 2 June 2021

Keywords:
philosophy of psychiatry; mental causation; top-down effects; control variable

Author for correspondence:
Kenneth S. Kendler,
E-mail: Kenneth.Kendler@vcuhealth.org

Abstract

Psychiatry has long debated whether the causes of mental illness can be better explained by reductionist or pluralistic accounts. Although the former relies on commonsense scientific bottom-up causal models, the latter (which typically include environmental, psychological, and/or socio-cultural risk factors) requires top-down causal processes often viewed with skepticism, especially by neuroscientists. We begin with four clinical vignettes which illustrate self-interventions wherein high-order psychological processes (e.g. religious beliefs or deep interpersonal commitments) appear to causally impact the risk for or the course of psychiatric/behavioral disorders. We then propose a model for how to understand this sort of top-down self-causation. Our model relies centrally on the concept of a control variable which, like a radio tuning dial, can implement a series of typically unknown physical processes to obtain the desired ends. We set this control variable in the context of an interventionist account of causation that assumes that a cause (C) produces an effect (E) when intervening on C (by manipulating it) is associated with a change in E. We extend this framework by arguing that certain psychological changes can result from individuals intervening on their own mental states and/or selection of environments. This in turn requires a conception of the self that contains mental capacities that are at least partially independent of one another. Although human beings cannot directly intervene on the neurobiological systems which instantiate risk for psychiatric illness, they can, via control variables at the psychological level, and/or by self-selection into protective environments, substantially alter their own risk.

One of the longest and most vigorous debates in the history of psychiatry is the degree to which mental illness can be best explained by reductionist strategies which seek the etiology of mental illness in various aspects of brain structure or function or by pluralist approaches (Mitchell, 2003) which assume that true causes of psychiatric disorders arise from a wide range of biological, psychological and social phenomena (Eisenberg, 1986; Kendler, 2005). One of the major appeals of reductionist approaches is their reliance on the commonsense scientific model of *bottom-up causation*. Take for example a Mendelian neurological disorder. It is intuitively easy to understand how an excess number of triplet repeats in a protein coding region of the genome could produce an abnormal protein. That protein then folds improperly thereby producing a toxic product that causes the degeneration of key neuronal populations. The cellular degeneration then leads to neurologic illnesses such as Huntington's chorea or Mendelian forms of Parkinson's or Alzheimer's disease.

Pluralistic accounts of the etiology of psychiatric illness cannot, however, rely entirely on such common-sense scientific models of bottom-up causation (Kendler, 2012a). Many well validated causes of psychiatric and substance use disorders include a range of risk factors acting at psychological and social levels such as stressful life events, childhood sexual abuse and peer influences (Kendler, 2014). These factors cannot easily be understood to impact risk only through bottom-up molecular pathways. Instead, to one degree or another, pathways to illness from many of these kinds of risk factors require some form of *top-down causation* – that is, causes at 'higher' levels than the biological. Although these models have intuitive appeal, they lack the simplicity of the neurobiological bottom-up causal models. Furthermore, these models cannot avoid the question of how, in a physical world, we can understand causation that originates at the mental level.

In this essay, the authors – a research psychiatrist (KSK) and academic philosopher with expertise in the problems of causation (JW) – take an approach opposite to that of the typical reductionist paradigm for psychiatric illness. Instead of molecular variants, we examine emergent events that either impact strongly on risk for psychiatric, behavioral or drug use disorders, or have large consequences on the course of these disorders. We illustrate such emergent events in four vignettes collected by KSK from his clinical and research experiences and readings, elaborated upon and modified to protect subject identity. These are described in Table 1. A key feature of all these vignettes is that the individuals themselves are initiating the causal

CAMBRIDGE UNIVERSITY PRESS

Table 1. Four clinical cases demonstrating self-initiated top-down causal effects

Case	Text
# 1	*Robert* was 52 years old when contacted to help us with a study of sibships with a high density of alcoholism. Two of his younger siblings, he was the oldest, had been treated for alcoholism and we were contacting him for help tracing them. He immediately volunteered his own story. Raised as a devout Catholic, he has been an abstainer all his life and watched, with horror, how alcoholism had consumed several of his younger brothers. His father had been a heavy drinker and things got worse when he was a small boy. Often, on Friday nights when he was to bring home his weekly pay to his mother who needed it to buy the family groceries on Saturday, he would come home late and very drunk having spent much of his pay check at the local bar. There was then yelling and sometime his father would hit his mother. During those terrible fights, Robert tried to comfort his younger siblings. Each time, he felt the hatred for this father and his drunkenness rising in him. Finally, when he was around 12, he couldn't stand it anymore and for the first time, he went into the kitchen to get between his parents and defend his mother. His father was in a rage and grabbed Robert, still far smaller than him, and threw him against the wall, going after his mother again. Robert recalled at that moment, praying to God. His prayers were something like 'Please God, please whatever happens to me never let me be like him, never like him'. Then it struck him that of course he never could drink, perhaps that would be the way to avoid 'becoming like him'. He never touched alcohol in his life.
# 2	*Jill* was 26 years old and had had a cocaine 'habit' for nearly 6 years. She had been in multiple therapies and short periods of abstinence, but always relapsed. 'It was so much fun' she would say. She had kept the habit – although increasingly expensive – under fairly good control. She was a talented and hard worker who rose up the ranks of a young start-up company so soon money was not a major problem. She met a co-worker there and they married. Her husband knew about her habit but not the extent of it. She got pregnant and managed to stay off the cocaine most of the pregnancy but relapsed again. Over the next 2 years, her habit accelerated and for the first time she got scared she would lose control. Therapy again produced only short remissions. More and more time was being taken up buying and using in secret, managing her funds so her husband wouldn't find out how much she was spending. Then, one morning before work, when her daughter was nearly two, she had forgotten to lock the bathroom door when she went in for her morning cocaine snort. Just as she was inhaling, her daughter toddled into the bathroom, looked very surprised and asked, 'What are you doing, Mommy?' Jill looked down. She saw the distress in her daughter's face. She said later 'It was like a switch. When I saw her looking up at me nearly in tears, clearly worried about her mother, I realized that was it. I could never do that again. I just couldn't. I loved her too much'. She stayed off cocaine over the next several years till we lost touch.
# 3	*Roseanne* and her monozygotic twin sister, Mary, grew up in a devout Roman Catholic household dominated by marital conflict. Their parents had repeated separations leading to a permanent separation when the girls were 12 years. Their mother had several mild depressive episodes in response to the earlier conflicts but then had a severe, impairing episode of major depression when the father announced he was leaving. Roseanne described herself as sensitive and nervous as an adolescent. She began dating at age 16 years and found herself easily upset by the ups and downs of teenage romance. She became increasingly involved in her religion and after graduating from high school became a nun. Her twin sister, Mary, married her high-school sweetheart, but that marriage broke up after 4 years and two children. Mary had repeated depressive episodes over her adult life, all related to unsuccessful romances. When we interviewed Roseanne, she was 52-years old and denied any history of mood or anxiety disorders. When asked to reflect on the differences in her life course and that of her identical cotwin, she said, 'I have often thought of this. Looking back, I think I realized I could not deal with men. I saw what was happening to my mother and saw the same future for myself. It was then that I decided it was better to fall in love with God. I have had a rich and wonderful life, and I feel so badly that Mary has had such an awful time of it. (Kendler, 2012b, p. 641)
# 4	*Charles* was 34 the youngest of a large family and soon became its 'black sheep'. His criminal career started young, initially carjacking. He dropped out of high school, had a spotty work record as an automobile mechanic and spent more time in than out of jail for much of his twenties. On his release, he would typically go back to his criminal activities quickly, eventually getting caught again and again. His parole officers despaired, declaring he had no 'rehabilitative potential'. He graduated to robbing liquor stores and in his late twenties was caught using a handgun so that resulted in his longest prison sentence. He never hurt anyone – something he was always proud about. His family eventually gave up on him. They refused to come to court, visit him in jail, all except his oldest sister Mary. 'She was kind of like a mother to me. She was the only one who ever really loved me. She would visit every Saturday without fail often bringing with her one of my nieces or nephews'. By his early 30s, Charles' parents were both dead, Mary was divorced, and the family had pretty much disintegrated. She had not been looking well and on one visit, Chuck asked about her health. For the first time ever, he saw his older sister break down in tears. She had disseminated breast cancer. 'Things look bad Chuck. I still have two of mine in school. The doctor thinks I might have a year. You'll be out by then. You are the only one I have left. Can I count on you to look after my kids? You've got to go straight now. I need to be able to count of you.' The guards who had gotten to know Charles well over the last 15 years escorted him out 6 months later joking with him that they expected to see him back soon. He moved in with his sister, helped her through her terminal illness, got a regular job working as a mechanic and cared for his nieces and nephews. He remarked 'Every once in a while, driving by a liquor store, I get this thrill in my stomach. Wouldn't it be fun to …. But then I think to myself 'Can't do that. I promised my sister''.

process – that is, they are 'changing themselves'. Then we provide a philosophical schema for how the effect of these high-level causes and the associated top-down causation can be understood. Our aim is not primarily to make the point that top-down mental causation is possible – a conclusion that many would agree with. Rather, our goal is to propose a more specific account of how top-down causation might be understood, one that shows this notion to be a coherent one and that connects it to influential current theorizing about causation. We also attempt to elucidate one particular variety of top-down causation, illustrated by our vignettes, in which subjects intervene on their own thoughts and emotions to change their behavior. This is what we understand to be involved in self-changing top-down causation.

From a clinical perspective, these four stories share one key feature in common. Some external event happened to each individual (Robert – getting thrown against the wall as a boy by his father; Jill – her daughter in the bathroom seeing her snorting cocaine; Roseann – her trouble with boys in adolescence; Charles – his sister's request to 'go straight') which resulted in a major internal change in them which in turn impacted the risk or course of their disorder. The stories differ in one important way. For Robert and Roseann, the event results in their making changes – Robert didn't drink, and Roseann became a nun – thereby avoiding romantic entanglements with men – that markedly reduced their risk of developing, respectively, alcoholism and major depression. For Jill and Charles, by contrast, the event broke a maladaptive behavior pattern that had previously been resistant to change and resulted in a firm resolution that produced remission of their chronic cocaine use and criminality, respectively. Common clinical (and indeed 'human') intuition suggests that these changes were causal. For each subject, there is a time series of mental and behavioral events, a cognitive change

https://doi.org/10.1017/S0033291721001811 Published online by Cambridge University Press

discontinuous with what preceded it and then major behavioral changes that are plausibly attributed to the cognitive change. Our concern, however, is not to conclusively establish that causation was demonstrated in these cases but rather with making sense of these episodes, on the natural assumption that causation is present. We claim that to do this we have to invoke something like 'top-down causation' because the causal effects of these events cannot be meaningfully understood by starting at a molecular level. On the contrary, the causal effects seem to have occurred in 'high psychological space' involving such ideas as self-concept, religious feelings and the changes arising from intimate interpersonal commitments including love. Although related to our concept of self-change through psychotherapy, these stories have a further oddity – the mind–brain system of each of these individuals 'changed itself'. To put that more technically, Robert, Jill, Roseann and Charles each intervened on themselves. What kind of system can do that?

Our point of departure is a view of causation that is increasingly influential both in the philosophical literature as well as statistical models of causal inference in the social and medical sciences. This is the *interventionist* account of causation (Kendler & Campbell, 2009; Woodward, 2003). The basic idea is simple. Cause (C) causes Effect (E) when if there is an appropriate intervention on C, E changes. A paradigmatic example of an intervention is an unconfounded experimental manipulation – a change in C that is such that any change in E that occurs can only occur through the change in C.

For example, if you absent-mindedly flip a switch on and off and the light changes accordingly, you likely have successfully intervened on the switch. Because of the random nature of your manipulations, it is unlikely that they are correlated with any confounding factor that might independently cause the light to go on and off – hence you can conclude that it is the switch that causes changes in the light. However, the interventionist account does not require that experimental manipulation actually occur; rather the idea is that causation is present if, were an intervention to be performed on C, E would change. We can sometimes reliably infer that this condition is met when we have purely 'observational' (non-experimental) data. Importantly, the interventionist account is applicable both to bottom-up and top-down causation, to causal effects that are assessed using biological tools (DNA sequence) and psychological or social processes, and to causal processes that are best conceptualised as happening in the brain and in the mind (Kendler & Campbell, 2009). Social and psychological processes have effects on mental illness if, were the former to be changed by interventions, the presence or course of the illness would change.

One way of thinking about interventions is that they put the variable intervened on – the candidate cause – entirely under the control of the intervention; the intervention removes or 'breaks' the influence of other possible confounding causal factors that might influence the effect. For example, in a randomized controlled trial of the efficacy of a drug in promoting recovery from an illness, the result of the random assignment to active treatment or placebo groups is that the influence of other factors besides the drug that might influence recovery is removed by the design of the experiment. In the light switch example, the position of the switch is (unless something very unlikely is going on) entirely under the control of the person manipulating the switch. In effect, by intervening we give the cause a new, independent and exogenous causal history – exogenous to the other 'endogenous' possible causal factors that might influence recovery. If the effect follows in such cases, we can be confident that it is attributable to the factor that is being manipulated. As we will argue below, this picture fits the self-intervention that we think is present in our examples.

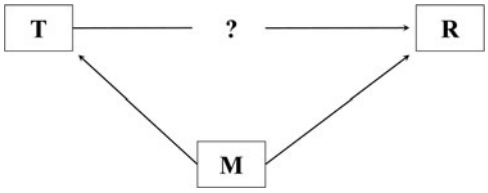

Fig. 1. The causal structure depicted here represents the case in which the possible effect of T on R with the arrow with the question mark representing that we are uncertain whether T causes R. This is what we want to determine.

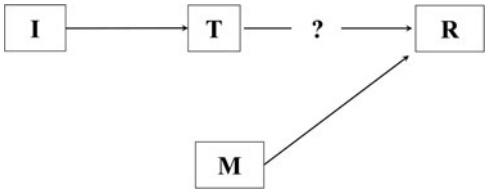

Fig. 2. The causal structure depicted here shows that the intervention I depicted here (e.g. the adoption of the randomized design) 'breaks' the connection from M to T, so that the value of T is determined just by the randomization procedure I. If, in this structure, treatment T is correlated with recovery, we can conclude that T causes R.

A standard device for representing this is in terms of directed graphs. When C directly causes E, this is represented by an arrow that is directed from C to E: C→ E. Suppose that recovery R from an illness is influenced by a patient's immune response M. If M influences both who is treated with a drug (the drug is given preferentially to patients with stronger immune response) then the effect of the drug on recovery is confounded with the influence of M. A randomized control trial in which subjects are randomly assigned to treatment and control groups can be thought of as removing any correlation between M and T (which measures who gets the treatment). Thus, the randomization, which can be thought of as an intervention I on T, replaces the structure see in Fig. 1 with the structure seen in Fig. 2.

In the example above, the adoption of a randomized experimental design involved completely 'breaking' the influence of M on T. This 'hard' intervention contrasts with another possibility in which the intervention on the candidate cause C is 'soft'. Here rather than completely breaking all other influences on C, the intervention I supplies C with an exogenous source of variation which is independent of the other endogenous causes of C. If this exogenous variation in C is associated with variation in E, one concludes that C causes E. We include this possibility because some psychological interventions are better modeled as soft interventions while others are more naturally modeled as hard.

Finally, a critical feature of an interventionist framework that we could apply to our clinical examples of top-down causation is that of a *control variable* (Campbell, 2010). A control variable is an interventionist cause with some critical additional properties. Most importantly, by employing appropriate control variables, we can bring about an effect regardless of the details of how that effect is actually implemented. For example, a professor might announce to her graduate research assistants 'We will meet at 1pm tomorrow in room 101'. Different students may have

https://doi.org/10.1017/S0033291721001811 Published online by Cambridge University Press

various reasons for following this directive: one because she enjoys the research, another because he wants to use some of the data for his thesis. The professor need not know or care about these details of student motivation (and still less about the details of how these are realised in the assistant's brains); making the announcement will bring about the desired effect (attendance) regardless. The announcement is, for the professor, a control variable for attendance.

Often, if one wants to influence another's behavior the best control variables are 'upper level' manipulations that act on mental states such as beliefs or desires. Consider this example:

> You and a friend are walking down the street deeply engaged in a conversation. You look down and see a large hole in the sidewalk right ahead of your friend. You say loudly 'Watch out. There is a hole in the sidewalk in front of you!' Your friend looks up and steps to the side avoiding the hole.

Your intervention supplied your friend with new information and started a very complex set of events from his ear to his auditory cortex, to his motor cortex, to a complex set of neural signals to his leg musculature etc. But you didn't need to know or even care about those details.

As an additional illustration, consider contingency management interventions which are among the most effective treatments for substance use disorders (Dutra et al., 2008). This treatment approach provides rewards, usually money, for drug-free urines which increases over time but is set back to zero for a 'dirty urine'. Contingency management, which we suggest can be considered an example of top-down causation, intervenes at a high level on the motivations for maintenance of drug abstinence. Designing and implementing this intervention requires no knowledge of the neurobiology of drug addiction, although, of course, the clients' decision to value the monetary incentives over further drug use is instantiated in neurobiological systems.

Although not our focus here, this idea of a control variable can be easily applied to conventional explanations for the efficacy of cognitive-behavioral therapy (CBT) for depression (Beck & Alford, 2008). (Although some of the efficacy of CBT is undoubtedly due to non-specific effects, there is good but not undisputed, evidence that some of its therapeutic benefit arises from the specific cognitive interventions; Honyashiki et al., 2014.) When a patient and therapist discuss a recent marital conflict, the patient interprets this example, like many others before it, to be entirely her fault, again demonstrating what a bad wife she is. When the therapist proposes a different interpretation, pointing out the unreasonable demands and irritability of the husband, she is hoping to intervene on a set of entrained cognitive processes whereby self-deprecating thoughts lead to a deepening negative self-evaluation and increased depressive symptoms. Our therapist need not know any details of how this all happens in the patient's brain. In fact, it is likely that such a change can occur through a myriad of different specific neural mechanisms in different patients (a feature philosophers' call 'multiple realizability').

If a therapist can intervene on a patient's mental states so as to alter behavior, it is but a small step to suggest that an individual may be able to do this herself, intervening on herself through the appropriate control variables. This is likely to be what happened in the cases with which we began this article.

The specific control variables utilized in psychiatrically relevant self-interventions can vary. One approach is to implement a decision to change the external environment – to move from a high risk to a low-risk environment. Roseanne provides us the clearest example of the former strategy. By joining a convent and following its dictates, she eliminated the high-risk environmental experiences of relationship difficulties with men that she perceived in her mother and sister. In terms of our discussion above, Roseanne's intervention breaks or disconnects a previously existing causal relationship between an environment in which she is exposed to romantic difficulties with men and a consequent risk for depression. She accomplishes this disconnection by moving to a different environment in which romantic relationships with men are not a possibility.

Robert's environmental intervention was even more specific. Perceiving that much of what he hated about this father was his drunkenness, he proposed a remarkably simple and effective intervention if maintained: avoid alcohol consumption. This intervention might have been of particular efficacy for Robert given his likely high genetic risk for alcoholism. However, Robert's story is far from unique. Rates of lifetime abstinence from alcohol are well known to increase in offspring of individuals with alcoholism (Harburg, Davis, & Caplan, 1982; Harburg, DiFranceisco, Webster, Gleiberman, & Schork, 1990). Although we will not pursue this further, we should note that religious beliefs played an important role in both Rosanne's and Robert's self-interventions.

Another approach to self-intervention would be to impact directly on one's own mental states and/or higher order self-concepts. Jill's decision to stop using cocaine has this character, although it's unclear how best to describe it. One approach might be to posit that she down-regulated her desire for the drug in comparison with her desire to have a satisfactory relationship with her daughter, although the feasibility of such down-regulation has been questioned (Berridge & Robinson, 2016). But that does not capture the salience and immediacy of her re-telling. Instead of considering this question at the level of up or down regulation of her drug desires in competition with her concerns for her daughter, we would favor invoking an intervention involving Jill's higher-order self-concept – that of being a good and loving mother. Since her daughter had been born, she uneasily balanced that self-concept with her cocaine use. However, that compromise collapsed when she saw the distress on her young daughter's face having caught her snorting cocaine. She couldn't do both and there was then no question of which self-concept – cocaine user or good mother – would win out. As we note below, this goal of maintaining a certain self-concept ('I don't want to be the kind of mother who…', 'I don't want to be a drunk like my father') seems to play an important role in the self-initiated top-down psychological changes present in several of our examples.

By comparison, the self-intervention with Charles, prompted by the request of his sister, fits within both levels: desires and higher order self-concepts. Particularly illuminating is his comment about the urges to rob while passing by a liquor store. Here, he acknowledges a re-emergence of his old desires, but this is promptly checked by his commitment to his sister which – in an adult life not generally characterized by strong connections with others – stands out for its strength and endurance.

Our four scenarios all differ on the degree of self-reflection and self-knowledge involved. Roseann stands at one extreme. Insight into the nature of the problems in her mother's and sister's relationship with men, and her own emerging problems were required. She needed to be able to project herself forward in time to consider a life like her mother's marked by marital strife and depression. Robert's intervention occurred suddenly with less

https://doi.org/10.1017/S0033291721001811 Published online by Cambridge University Press

self-reflection and in the setting of strong emotions, although as noted above, it was very effective. Jill's intervention was also sudden and involved little to no self-reflection. Unlike the others, Charles's self-intervention was precipitated by an external request. But as best we understand, not much self-reflection was involved at least at first. Of course, he would do what his sister asked of him. She had always stood by him. How could he do less?

Conceptually, the idea of voluntarily changing one's external environment, as Roseanne did, seems unproblematic. However, the idea that one can voluntarily change or intervene on one's own mental states might seem more questionable. Indeed, a number of philosophers claim that for conceptual reasons one cannot voluntarily change one's beliefs via anything that looks like self-intervention (Williams, 1970). Genuine beliefs, they argue, can only change in response to external evidence and the impact of evidence on belief is outside the control of the believer. In addition, a common view of desires (held not just by philosophers but by many economists) is that our fundamental desires or preferences are typically fixed in strength and outside of voluntary control. Given such a picture, when we act, our actions simply reflect whatever desire happens to be strongest. It is as though the self is just a passive spectator that observes its behavior unfold as a consequence of its beliefs and desires. The self can't intervene to *do* anything – or at least nothing over and above what its present belief and desires 'do'. (Historically views like this have been held by some leading historical figures in philosophy such as Hobbes and Hume and are assumed in many philosophical discussions of 'belief/desire psychology'.)

We think that this is an oversimplification. First, many beliefs or belief-like states have an 'active', volitional or commitment component – they are not just passive states induced by one's environment. This is particularly true of beliefs about oneself, as well as moral, political and religious beliefs. In these areas, belief often involves an element of active commitment or endorsement by the self – one actively 'identifies' with the belief or thinks of it as part of one'sself -identity. This would explain part of why Jill could change her beliefs about her cocaine habit because she could not be the 'kind of mother' who would expose her daughter to her cocaine use.

The efficacy of CBT in the treatment of depression provides evidence for our ability to actively modify our beliefs – in this case fundamental, although often irrational, views about our self (Cuijpers et al., 2013). Some degree of voluntary control over one's desire or evaluations also seems possible as illustrated in the cases of Jill and Charles. Modulating desires through changes in attentional focus (something that is somewhat under voluntary control) is one possible strategy for accomplishing this, as when, seeking a dessert at a luncheon buffet, you stop looking at the cookies, and direct your attention instead to the fruit bowl.

To connect these descriptions with earlier parts of our discussion, recall that which is required for an intervention on some factor C with respect to a second factor E is that the intervention must involve an independent or exogenous change in C, that either disconnects other causal influences on C (a hard intervention) or else supplies C with an independent source of variation (a soft intervention). One might think of Jill's decision as either disconnecting the influence on her behavior of her previously existing endogenous desire for cocaine or as reflecting the influence of motivations (related to the sort of person that she wants to be) that operate as soft interventions in the sense of being exogenous to (and independent of) her desire for cocaine. In the latter case, her desire for cocaine is not disconnected but an additional motivation kicks in that operates independently of her desire for the drug.

It should be clear from our discussion that this notion of intervening on one's own mental state will make sense if there are independent psychological mechanisms and associated control variables that can potentially be brought under voluntary control and that can act on one's previously existing ('endogenous') beliefs or desires or other mental states, altering them or removing or modulating their influence and in this way influencing behavior. This requires what might be described as a non-unitary view of the self: the self has different capacities or states that can operate at least somewhat independently of each other. In particular, for self-intervention on one's mental states to be possible, there must be control mechanisms/variables that can operate independently or exogenously on some of a subject's beliefs, desires and other mental states. For example, we sometimes seem able to alter attentional focus in way that is not caused or determined by our desire for cookies – our attentional focus is sometimes under independent control and we can use this to influence the strength of our desire for cookies or whether it influences our behavior.

We want to emphasize that we don't intend anything mysterious or magical by this. Presumably attentional focus and decisions about its allocation have ordinary physical causes or realizations just like everything else in the mind/brain. What is crucial is that the allocation of attention can be intervention-like – not uncaused but rather reflecting the operation of causes or mechanisms that can function independently of the desires that they influence. In the same way, the behavior of a scientist who implements a randomized experiment is not uncaused or outside the realm of ordinary causation; what is crucial is rather that the allocation to treatment and control groups be exogenous to or independent of other possible causes of the effect of interest. Similarly, Roseanne's intervention to change her environment presumably had ordinary psychological causes; what is important is that these causes were independent enough of the factors influencing her relationships with men that she could choose a new environment that disconnected the relationship between those factors and depression.

We argued above that appropriate control variables for influencing behavior are often at the psychological level. This is also true for self-interventions. We cannot directly control which ionic currents are present in particular neurons in our brains. We can however influence our brains and minds via self-interventions at the psychological level – by operating on our beliefs, desires and other mental states. It is conceptualizations at this level that provide 'handles' for self-intervention in the examples discussed above and that allow subjects to envision possible reconfigurations of their mental states and environments that might be desirable and so to reason counterfactually about these. (If I were to direct my attention away from the cookies, then…) Human beings typically have no clue about how they might implement these sorts of reconfigurations by acting directly on their brains – they lack access to the relevant control variables. This is likely also true for neuroscientists at the present time. This is one of many reasons why conceptualizations at the psychological level are likely to remain part of the discipline of psychiatry – in many cases we need such conceptualizations if we are to act on both our own mental states and those of our patients in the manner illustrated by our examples.

We said above that we can sometimes control our beliefs and desires but that we cannot control, in the same way, the detailed behavior of the brain states that realize these. Our picture is this: When I intervene to modify one of my desires, some realizing

https://doi.org/10.1017/S0033291721001811 Published online by Cambridge University Press

brain state must of course be present but there are a number of different brain states that can realize my desire (multiple realization) and I don't control which particular brain state occurs as the realizer of my desire. In the same way, if I were to place a sample of gas in a heat bath, this would be an intervention that controls its temperature, but the particular constellation of molecular kinetic energies that realizes that temperature value is not something that is controlled by the heat bath intervention and indeed will vary from moment to moment. The person who wishes to control her behavior is in the same position as the researcher who wishes to control the upper level state of the gas. The researcher has an available upper-level intervention (on temperature) for doing this but, in any ordinary situation no procedure for intervening on individual molecules. Similarly, the person who wishes to modify her behavior will often only have available possible interventions on her upper-level mental states.

A final remark: our analysis suggests that those disorders for which self-intervention may be especially efficacious are those which have a volitional component. These are most clearly illustrated by disorders involving substance abuse, antisocial behaviors or eating disorders. But our story of Roseann suggests that such an approach may also apply to disorders for which, given sufficient prescience, high-risk environments could be avoided.

Acknowledgements. Peter Zachar PhD provided helpful comments on an earlier version of this essay.

Conflict of interest. None.

References

Beck, A. T., & Alford, B. A. (2008). *Depression: Causes and treatment* (2nd ed.). State College, PA: University of Pennsylvania Press.

Berridge, K. C., & Robinson, T. E. (2016). Liking, wanting, and the incentive-sensitization theory of addiction. *American Psychologist*, *71*(8), 670.

Campbell, J. (2010). II – control variables and mental causation. Paper presented at the Proceedings of the Aristotelian Society.

Cuijpers, P., Berking, M., Andersson, G., Quigley, L., Kleiboer, A., & Dobson, K. S. (2013). A meta-analysis of cognitive-behavioural therapy for adult depression, alone and in comparison with other treatments. *The Canadian Journal of Psychiatry*, *58*(7), 376–385.

Dutra, L., Stathopoulou, G., Basden, S. L., Leyro, T. M., Powers, M. B., & Otto, M. W. (2008). A meta-analytic review of psychosocial interventions for substance use disorders. *American Journal of Psychiatry*, *165*(2), 179–187. doi: appi.ajp.2007.06111851 [pii];10.1176/appi.ajp.2007.06111851 [doi].

Eisenberg, L. (1986). Mindlessness and brainlessness in psychiatry. *The British Journal of Psychiatry*, *148*(5), 497–508.

Harburg, E., Davis, D. R., & Caplan, R. (1982). Parent and offspring alcohol use; imitative and aversive transmission. *Journal of Studies on Alcohol*, *43*(5), 497–516. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/7144182.

Harburg, E., DiFranceisco, W., Webster, D. W., Gleiberman, L., & Schork, A. (1990). Familial transmission of alcohol use: II. Imitation of and aversion to parent drinking (1960) by adult offspring (1977) – Tecumseh, Michigan. *Journal of Studies on Alcohol*, *51*(3), 245–256. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/2342364.

Honyashiki, M., Furukawa, T. A., Noma, H., Tanaka, S., Chen, P., Ichikawa, K., … Caldwell, D. M. (2014). Specificity of CBT for depression: A contribution from multiple treatments meta-analyses. *Cognitive Therapy and Research*, *38*(3), 249–260.

Kendler, K. S. (2005). Toward a philosophical structure for psychiatry. *American Journal of Psychiatry*, *163*(3), 433–440.

Kendler, K. S. (2012a). The dappled nature of causes of psychiatric illness: Replacing the organic-functional/hardware-software dichotomy with empirically based pluralism. *Molecular Psychiatry*, *17*(4), 377–388. doi: 10.1038/mp.2011.182

Kendler, K. S. (2012b). Decision making in the pathway from genes to psychiatric and substance use disorders. *Molecular Psychiatry*, *18*(6), 640–645. doi: mp2012151 [pii];10.1038/mp.2012.151 [doi].

Kendler, K. S. (2014). The structure of psychiatric science. *American Journal of Psychiatry*, *171*(9), 931–938. doi: 1873859 [pii];10.1176/appi.ajp.2014.13111539 [doi].

Kendler, K. S., & Campbell, J. (2009). Interventionist causal models in psychiatry: Repositioning the mind-body problem. *Psychological Medicine*, *39*(6), 881–887. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/18845010.

Mitchell, S. D. (2003). *Biological complexity and integrative pluralism*. Cambridge: Cambridge University Press.

Williams, B. (1970). Deciding to believe. In H. E. Kiefer & M. K. Munitz (Eds.), *Language, belief, and metaphysics* (pp. 95–111). Albany: SUNY Press.

Woodward, J. (2003). *Making things happen* (*P. Humphreys Ed.*). New York: Oxford University Press.