HHS Public Access

Author manuscript

Psychol Inq. Author manuscript; available in PMC 2016 March 28.

Published in final edited form as:

Psychol Inq. 2015 ; 26(3): 209–230. doi:10.1080/1047840X.2015.1015367.

The Power of Theory, Research Design, and Transdisciplinary Integration in Moving Psychopathology Forward

Uma Vaidyanathan[a], **Scott I Vrieze**[b], and **William G. Iacono**[a]

Uma Vaidyanathan: vaidy017@umn.edu; Scott I Vrieze: scott.vrieze@colorado.edu; William G. Iacono: wiacono@umn.edu

[a]Department of Psychology, University of Minnesota, N218 Elliot Hall, 75 East River Road, Minneapolis, MN 55455

[b]Institute for Behavioral Genetics, University of Colorado – Boulder, 1480 30th Street, Boulder, CO 80303

Abstract

While the past few decades have seen much work in psychopathology research that has yielded provocative insights, relatively little progress has been made in understanding the etiology of mental disorders. We contend that this is due to an overreliance on statistics and technology with insufficient attention to adequacy of experimental design, a lack of integration of data across various domains of research, and testing of theoretical models using relatively weak study designs. We provide a conceptual discussion of these issues and follow with a concrete demonstration of our proposed solution. Using two different disorders – depression and substance use – as examples, we illustrate how we can evaluate competing theories regarding their etiology by integrating information from various domains including latent variable models, neurobiology, and quasi-experimental data such as twin and adoption studies, rather than relying on any single methodology alone. More broadly, we discuss the extent to which such integrative thinking allows for inferences about the etiology of mental disorders, rather than focusing on descriptive correlates alone. Greater scientific insight will require stringent tests of competing theories and a deeper conceptual understanding of the advantages and pitfalls of methodologies and criteria we use in our studies.

Keywords

Etiology of mental disorders; Neurobiology; Pitfalls and advantages of technology and statistical modeling; Quasi-experimental designs; Theory vs data-driven approaches

---

Not everything that can be counted counts, and not everything that counts can be counted

- (Cameron, 1963)

1. Introduction

Understanding the etiology and structure of mental disorders is the purpose of psychopathology research, and an especially timely issue presently with recent revisions to the Diagnostic and Statistical Manual of Mental Disorders (DSM), and initiatives such as the Research Domain Criteria (RDoC). Evidence from several different lines of work such as

behavioral and molecular genetics, neuroimaging, family studies, etc. has been considered in both of these initiatives but transformative progress has been more difficult than anticipated (Kupfer & Regier, 2011). We contend what is particularly required is a framework and strategy for integrating the data from such diverse studies. Cronbach (1957) proposed that studies in psychology could be broadly divided into two streams – experimental and correlational. Such a strict division may not necessarily be the case in current times, and a range of studies exist that amalgamate these two streams of thought to differing degrees, such as quasi-experimental research (Rutter, 2007; Shadish, Cook, & Campbell, 2001). Indeed, this kind of paradigm – one that combines the causal significance of laboratory tests with the import of studies involving people as they function in the real world – is much needed in psychopathology research and should be advocated on a much grander scale.

The place to begin is with a guiding theory that informs how to evaluate evidence for or against a particular hypothesis. Theories may be derived from pre-existing notions of a particular disorder, or may arise out of the data after evaluating a certain amount of evidence for or against it (such as with exploratory data analysis; Tukey, 1980). The strategy we advocate here, however goes beyond that – here, we focus on the interface between research design, statistics, psychology, and neurobiology, and the extent of theoretical inference that can legitimately be drawn from them. Such an approach will require us to (a) move beyond an overreliance on advanced statistics and the use of new technology, especially in the absence of causally informative research design; and (b) integrate data in a meaningful way across methodologies using theoretically guided designs, and (c) test such theories using stringent study designs that allow us to make less ambiguous inferences regarding causality. In particular, the emphasis on the process of quantification and the use of advanced technology in some research occurs at the expense of theoretical or study design considerations, and lends an illusory quality of validity or deeper meaning to the results than warranted (Tavris, 2010) – a process that can easily lead to misguided conclusions and misallocation of research dollars.

Stated more concisely, to advance our understanding of the etiology of mental disorder, we as a field must employ, whenever possible, risky tests of causal theories. What is a risky test? The risky test, in our view, is or approximates an experimental design and narrows the number of interpretive possibilities regarding causality, such that the riskier the test, the stronger the inference one can draw about the theory being tested. The ultimate risky test provides the opportunity to falsify and reject a theory, but we recognize this is seldom if ever possible in our field (Meehl, 1990), in part due to the inherent limitations of true experiments in psychology, especially when dealing with the ethical and legal obligations involved in research with human subjects (Meehl, 1978). In the absence of a true experiment, there are however, many clever ways to devise tests to be as risky as possible. Thus, we will use the phrase “risky test” to refer to (quasi-) experimental designs that not only allow strong inference about the tested theory, but are also practically feasible. If we perform risky tests from a variety of independent study designs then, when the results converge on the same interpretation, we have especially compelling evidence for theory corroboration or falsification.

We turn now to a conceptual discussion about the utility of statistics in testing theories about psychopathology, the use of novel technology in understanding mental disorders, the relationship between these domains and, lastly, the possible inferences one can legitimately draw this research. Our point here is not to examine the philosophical bases underlying mental disorders or statistics, though we do occasionally cover such concerns as part of our discussion. Nor do we focus solely on statistical issues such as replication, or multiple testing, or even questionable research practices (for an excellent discussion on these and related topics, see November 2012 issue of Perspectives on Psychological Science Pashler & Wagenmakers, 2012). Rather, the crux of our paper is focused on how to simultaneously leverage all our available research tools, use them to test etiological theory, and thereby arrive at the causal insights necessary to create, for example, a psychiatric nosology grounded in etiology, rather than one centered on descriptive psychopathology. We follow this with a concrete demonstration of our proposed solutions using two different disorders – depression and substance abuse – and link them back to the conceptual issues we have highlighted.

2. Conceptual issues on the use of statistics and technology

2.1 Statistics have their limits

Given our strong focus on the scientific process and strategies for deriving meaning from the results it provides, a few working definitions are useful. On pain of simplicity, we adopt a Popperian philosophy of science where the scientific process is a cycle in which one outlines a theory (a set of hypotheses based on prior data), designs and runs an experiment to test it, evaluates the findings, and accepts, rejects, or modifies the theory based on the results, runs another experiment if necessary, and so on (Lakatos, 1978; Meehl, 1990; Popper, 1959). There are several philosophical considerations here as to what it means to “accept” or “reject” a theory, but at least this is the basic premise. One option would be to ignore the theory part and rely solely on statistics, technology, and data mining. However, though some amount of tinkering and exploratory analyses can be useful and important (Behrens, 1997; Tukey, 1977), blind reliance on such methods cannot be counted on to provide consistent fuel for scientific progress or to aid the development of a larger theory of mental illness, no matter how “big” the data are. We shall discuss some examples below that illustrate why. A second option is to rely exclusively on theory to guide all our research; however, such a strategy would be focused solely on justification of existing theories, rather than allowing data exploration to spark new ideas and raise new questions. Another choice is to develop and refine theories in conjunction with clever research designs and risky tests. This may seem like a painfully obvious point to most readers but, as will be seen in the following sections, what may be obvious is not always implemented in practice.

First though, we will tackle the issue of what it means to test a theory in general. In contemporary psychology research, statistics are often seen as very useful ways to evaluate some theory or hypothesis. Most common statistics and statistical models that we use are based on linear relationships among variables, such as a correlation. Regardless of the available data, whether they are personality questionnaires, clinical interviews, electroencephalographic (EEG) brain oscillations, MRI scans, or genetic polymorphisms, it

remains true that a correlation, analysis of variance, regression, principal component analysis, factor analysis, structural equation model, or any other approach, evaluates the degree to which a set of variables behave in relation to each other. In other words, they give us an idea of the kinds of lines we can fit to the dataset at hand – straight, curved, going up, down, etc. By themselves, they give little to no information about *why* the variables go up and down together, just that they do. The additional catch here is that there are an infinite number of lines (literally) that can fit any dataset. While we use the line of "best" fit as a rule of thumb, and generally use straight lines in our statistics, there are no limits to lines we can use (Breiman, 2001b; McDonald, 2010; Meehl, 2002).

A justified response from many scientists to the issue raised above (i.e., choosing the best-fitting line) is to use the principle of parsimony to choose the best model. This may sound easy, but quantifying parsimony is not a trivial issue with clear guidelines. For example, Meehl (2002) identified no less than four common conceptions of parsimony, including curve fitting, economy of theoretical postulates, economy of theoretical concepts, and Ockham's Razor, and these were described in the context of 11 total criteria that scientists use in appraising scientific theories. It is far from clear why one particular definition of parsimony is the most important when it comes to deciding upon any one particular theory. However, the consequences can be striking depending on the choice (e.g., deciding whether the latent structure of psychopathology is categorical or continuous based on parsimony as the simplest line fitting the data using fit indices such as the Akaike Information Criterion or Bayesian Information Criterion; cf. Grove & Vrieze, 2010).

Just as important as quantifying model fit is the interpretive step from selected model to scientific take-home message. Unless one has selected a true model, the very fact that individual models are selected and interpreted means that bias is incurred in the resulting parameter estimates. Drawing conclusions (e.g., about the etiology and structure of psychopathology) as if the selected model is the only model to be considered ignores the fact that any selected model, no matter the weight of evidence in its favor, may have been selected in error. Thankfully this issue, referred to in the literature as model selection uncertainty, is becoming increasingly recognized as a major issue in statistical inference (Sterba & Pek, 2012; Yuan & Yang, 2005). Related to this, model averaging approaches (Claeskens & Hjort, 2008), where perhaps thousands of models are fit to a dataset through resampling methods, show that using all of the models is often predictively superior to methods where only the best model is selected (Breiman, 1996, 2001a, 2001b; Yang, 2005; Yuan & Yang, 2005). These results are found both in simulation studies and real-world applications, such as making predictions about future events and waiting for those events to unfold. Termed *ensemble learning* or *ensemble classifiers*, the superior performance of these methods drive home forcefully the fact that our model selection procedures are probabilistic and error-prone, and that even the best model is inferior to extremely complex model averages (which seem very far from parsimonious). The drawback to ensemble classifiers is that it is difficult to draw simple and substantive scientific conclusions from them; on the other hand, the drawback to selecting a single best model is that it is overly simplistic, and investigators run the risk of overstating results and drawing unwarranted simple conclusions about complex phenomena. While simple models based on the notion of parsimony can

certainly be useful, parsimony itself is not a universal law, just a heuristic. Indeed, anti-razors such as Hickam's dictum ("Patients can have as many diseases as they damn well please") have been fruitfully used in other fields such as medicine (Abramowitz et al., 2008). Thus, one must be careful when emphasizing simplicity or parsimony, especially when model fitting is conducted in observational datasets, a best model selected, and inferences made from that model about the etiological structure of a set of variables.

What is required, then, to adequately evaluate a scientific question? Simple models can help make sense of unruly data, but heavy reliance on statistics is a poor route to scientific insight. As everyone reading this knows, it takes a question, a theory, a design, an analysis, and a result. The difficulty comes when actually implementing any of these steps. The theory is most useful when it postulates a risky hypothesis, preferably including some etiological mechanism, although strong theories are often elusive in psychopathology research. The design must then provide a test of the etiological hypothesis which, for most interesting questions, requires an experimental or quasi-experimental design. Statistics is clearly crucial to the analysis step, with the correct statistical approach depending on the question, theory, and design. Finally, conclusions are something with which statisticians can assist, but depend just as strongly on the scientific adequacy of the experimental design to measure and control variables relevant to the etiological hypothesis. We address each of these issues below, starting with an important distinction between descriptive and etiological hypotheses.

2.2 Testing a theory rigorously requires a risky experimental design

The more fundamental point behind the issue raised in the section above is whether we can use our knowledge of statistics and modeling to obtain insight about the etiology of mental disorders versus a description of how the disorders relate to one another. This requires etiological theories, *and* the submission of those theories to risky tests (Meehl, 1978, 1990; Popper, 1959). As mentioned earlier, risky tests refer to practical research designs that narrow the number of interpretive possibilities for our results. When two variables correlate it is reasonable to infer that they somehow share space in a common causal framework (Cronbach & Meehl, 1955). Despite what we teach undergraduate students, correlation implies causation. The problem is that it is very difficult to make inferences about a causal framework from correlational data other than that a framework exists. What is more, deriving from some theory the prediction that two variables are causally related, then testing the theory by taking measurements and computing a correlation, is far from a risky test of theory; the test can easily be passed even if the causal theory is false. An overreliance on statistical modeling separated from causally-informative research design and plausible etiological theory will never produce insight that goes beyond simple conclusions that two variables somehow share etiology because they correlate (Rubin, 2008). Taleb (2010) had an excellent analogy for this where he noted that fitting models to data and inferring causality is like figuring out the shape of a melted ice cube from the puddle of water that it has left behind. There are any number of ice cubes with varying shapes and structures that can generate any particular puddle of water.

Obtaining etiological insight (knowing the ice cube) requires etiologically informative research designs. Correlational data, also called descriptive or narrative data, whether longitudinal or cross-sectional often provide little leverage to address etiological questions because the *Fundamental Problem of Causal Inference* remains (Holland, 1986). This is the notion that to make causal inferences we would ideally experimentally manipulate an independent variable. In a longitudinal sample this would amount to observing Subject X have an experience (e.g., use alcohol excessively during adolescence) and not have an experience (e.g., abstain from alcohol) simultaneously, or to ocurr in parallel universes, which is probably not possible. The practical approach is typically to randomly assign participants to each condition. If this were possible – which it often is not – then we could rigorously test causal hypotheses, such as whether excessive alcohol use causes brain-related changes in the mesolimbic dopamine system or the prefrontal cortex, as some have hypothesized (Goldstein & Volkow, 2002; Kalivas & Volkow, 2005). If we only have longitudinal but otherwise etiologically uninformative samples, then we cannot observe every individual in their real life where they are exposed to the variable of interest as well as counterfactual life where they are not. We can, with a little ingenuity however, approximate a true experiment and get some information about the counterfactual state of affairs (Rubin, 2001, 2007, 2008). Monozygotic twins discordant for some variable (e.g., alcohol use) give one approximation to an experiment, as they provide for each other a quasi-experimental control for genetic and shared environmental background (McGue, Osler, & Christensen, 2010). The Minnesota Study of Twins Reared Apart is a discordant twin design (discordant for rearing environment) that provided a strong test of the causal effect of rearing environment on a wide range of psychological outcomes (Bouchard, Lykken, Mcgue, Segal, & Tellegen, 1990). This study was especially powerful because prominent psychological theory, such as psychodynamic theory, made predictions about the importance of rearing environment on a wide range of outcomes. The use of twins (Galton, 1875), children of twins (D'Onofrio et al., 2003), adoption designs (Rutter, 1998), natural disasters (Kilpatrick et al., 2007), migration studies (Ravussin, Valencia, Esparza, Bennett, & Schulz, 1994), Mendelian randomization (Ebrahim & Smith, 2008; Smith, 2011), and other *natural experiments* (Rutter, 2007), provide powerful methods to conduct quasi-experimental research (Shadish, et al., 2001) in humans when ethical and other considerations often prohibit true experimentation (Meehl, 1978). This is not to say observational research is not valid – indeed, one needs to start at descriptive research to even form a theory of some sort. However, the inferences that can be drawn from purely observational work are open to multiple causal interpretations, and thus, less risky than counterfactual designs such as the ones listed above, especially when it comes to positing etiological theories about mental disorders. However, it is worth emphasizing that, just as there are no surefire statistical methods for establishing the etiology of any disorder, neither are there any surefire research designs.

2.3 Advanced technology and complex statistical models do not always provide more insight

Recent years have seen exciting advances in our ability to use increasingly complex and computationally demanding statistical methods (e.g., latent trait, class, and factor mixture models) and technology that probes deeper and deeper into the human body (e.g., EEG to

MRI to molecular genetics) in an attempt to understand psychopathology. The issue here is not with the tools we use, but rather what can be a noncritical reliance on them, simply because they appear to be "advanced" in some way. For example, the explosion of structural equation modeling since LISREL (Jöreskog & Sörbom, 1986), and more recently Mplus (Muthén & Muthén, 1998-2010), allows researchers to very easily fit extremely complex models and interpret them, in some cases with nothing beyond a superficial understanding of the method. Conclusions from these models at times go far beyond what is warranted, such as that the latent entities exist in nature (Maraun, 1996; Meehl, 1992, 1993), or that a causal relationship is "suggested" because an arrow in the path diagram is asterisked at $p < .0001$. There is undoubtedly much to learn about psychological phenomenology from factor analyzing covariance matrices, or using other dimensionality reduction techniques (e.g. The Big Five (Digman, 1990), Internalizing/Externalizing (Achenbach, 1991; Krueger, 1999), *g* (Spearman, 1904)). What makes these models particularly useful is the fact that they summarize large amounts of data in a convenient way (e.g., all the personality descriptors in the English language can be condensed into three to five dimensions) and they give researchers a common parlance in which to discuss constructs. More broadly, such analytic techniques are particularly useful in inductive or exploratory contexts where the goal is dimension reduction to explore, organize, and summarize information. However, great caution is warranted before drawing conclusions, especially about etiology, on the basis of these studies alone. In other words, deducing that the Big Five Traits exist in reality or that Internalizing and Externalizing are the only way to organize mental disorders is a much riskier process.

This issue of overreliance on advanced models or technology is by no means restricted to statistical modeling. For example, in neurobiology, functional magnetic resonance imaging (fMRI) research is published frequently in prominent journals with inferences made about particular brain regions implicated or even causing certain disorders that appear to outstrip the research design in which the technology is used (e.g., research on the "neurobiological basis" of some disorder or emotion). Recent times, however, have seen a burgeoning critique of how these technologies are used to advance our field. For example, blood-oxygen-level-dependent (BOLD) contrasts are the primary variable of interest in fMRI analyses. Logothetis (2008), though, observed that inferring exactly what kind of activity a BOLD signal represents can be unclear –e.g., excitation vs inhibition, activity relative to what kind of baseline, etc. Likewise, in another widely discussed paper, Vul, Harris, Winkielman, and Pashler (2009) have pointed out that researchers can get artificially high correlations between brain activation and some personality measures by only using voxels that cross a certain threshold in their analyses. Another concern is that of sample size – while it is relatively easy to administer and analyze self-report data from thousands of subjects, this is not the case with neurobiological research for practical and cost-prohibitive reasons. Thus, most fMRI studies are restricted to smaller samples (e.g., less than 50), and draw conclusions about relative activity or inactivity in various brain regions based on such samples. These limitations can pose substantive problems for researchers in this area. For example, stuides (Gonzalez-Castillo et al., 2012; Thyreau et al., 2012) have found that when samples include large numbers of subjects or trials (> 500 or 1000), most regions of the brain show up as statistically significant in analyses. In fact, prior work has shown that the

average statistical power of neuroscience studies is quite low (Button et al., 2013). Of course, the issue that is being highlighted in these studies – statistical versus biological significance – is not specific to fMRI (Farah, 2014). But the point remains that it is unclear as to what exactly is the required evidence for drawing appropriate conclusions about brain regions implicated in mental disorders. Similar problems are present in other methodologies as well. Electroencephalography (EEG), for instance, has fine temporal resoluation (MRI does not) but it is a two-dimensional representation of a three-dimensional brain, and there is no way currently to identify a single anatomical source for any observed waveform. Consequently, knowing that two groups of subjects differed on some event-related potential (ERP) is not very informative unless we have a strong theoretical notion about the significance of that ERP in some experimental context.

Genomics represents another major technological innovation. Advances in this field have made measurement of millions of genetic variants in humans easy and inexpensive. The completion of the Human Genome Project was a landmark event in the history of science, and has had massive implications for biology that have gone far beyond the investigation of genetic etiology of disease (Lander, 2011). We have gained a great deal of knowledge about the genetic architecture of disease, and we now know that existing candidate gene studies for common psychiatric disorders and psychological traits based on a priori hypotheses about gene function are unreliable (Collins, Kim, Sklar, O'Donovan, & Sullivan, 2012), and have returned largely false positive results. This is in part due to the small effect of common single nucleotide polymorphisms (H. L. Allen et al., 2010; Schizophrenia Working Group of the Psychiatric Genomics Consortium, 2014; Speliotes et al., 2010; Sullivan, Daly, & O'Donovan, 2012) and the underpowered samples used to test their effect on disease (L. E. Duncan & Keller, 2011; Vrieze, Iacono, & McGue, 2012). The best single example is the serotonin transporter, initially thought to show interesting gene by environment interactions in predisposing to depression (Caspi et al., 2003). Though based on a plausible theory – i.e., that the serotonin system is implicated in depression – after a decade of research the initial findings remain inconclusive (Karg, Burmeister, Shedden, & Sen, 2011; Munafo, Durrant, Lewis, & Flint, 2009; Risch, 2009). The point: A more technologically sophisticated tool has not necessarily advanced our understanding of mental illness – it may in fact derail it for a time.

Fortunately for genetics, we now have genome-wide association studies (GWAS) and widely-accepted conventions to avoid false-positive results, which are now providing insight into the failure of prior candidate gene and linkage studies of complex traits/diseases (Visscher, Brown, McCarthy, & Yang, 2012). GWAS is predicated on a specific theory of genetic etiology, the common-variant common-disease hypothesis (Pritchard & Cox, 2002), and continues to make advances in understanding the genetic architecture of psychiatric disease (Sullivan, et al., 2012) but has its own limitations. While GWAS has been very successful for non-psychiatric complex diseases, the single most important result for us can be characterized as a very risky test of a large number of candidate gene hypotheses. GWAS has told us that even our most promising candidate gene theories were little better than random guesses at genotype-phenotype associations. For example, even in samples traditionally considered large (e.g., N = 10,000), serotonin-related genes show no more of a signal for depression than other randomly selected regions in the genome. It is on this basis

that many candidate gene theories in psychiatric genetics have been very strongly discorroborated. In fact, a recent laudable paper by Hart et al. (2013) that attempted to replicate the authors' own prior candidate gene findings showed that none of their 12 prior significant results held up even when attempting to replicate and extend their sample size to just 200 additional subjects. Now, we must keep in mind that such tests are not complete. The candidate gene theories could be revised to state that the effects of the candidate genes are just much, much smaller than previously expected (e.g., DRD2 and Schizophrenia; Schizophrenia Working Group of the Psychiatric Genomics Consortium, 2014). In addition, other hypotheses about genetic etiology such as the rare-variant common-disease hypothesis (Pritchard, 2001) are now capable of being tested with new technologies such as exome (Fu et al., 2013; Nelson et al., 2012) and whole genome sequencing (Abecasis et al., 2012), and we can expect scientific findings to continue to grow in biology and psychiatric genetics. Indeed, it may turn out that serotonin-related genes do have large effects, but we were unable to measure those effects until now because the relevant variants are very rare and heretofore unmeasured. We must be careful however that these new advances in technology are applied with appropriate care and circumspection in psychology and psychiatry.

Leveraging new technologies is crucial to the advancement of science, and always has been. The problem arises when technology is used without considering research design or substantive scientific theory. Carol Tavris (Voss, 2012) has referred to this tendency as "technomyopia" – or the tendency to be dazzled by the technology at hand rather than evaluating if what it reveals is substantive. Others (Marcus, 2012; O'Connor, Rees, & Joffe, 2012) have also noted that this tendency can become even more extreme when mainstream media pick up on such articles and try to convey them to readers. Part of our responsibility as a field involves "de-hyping" our results and conveying them in the proper context and with appropriate caveats. As a general rule, statistical models of correlations do not, without some kind of experimental control, allow etiological insight beyond that obtained with measures of association. Likewise, merely increasing the resolution of our tools or our sample size does not guarantee the worth of a study. Even the largest sample combined with the most advanced data mining techniques and the most sophisticated technology is only mildly informative for understanding etiology without some kind of (quasi) experimental control and, something we take up later, a good scientific theory that can account for results.

2.4 Relying solely on any single methodology can lead to incorrect conclusions

Descriptive, experimental, quasi-experimental: these and other diverse designs have strengths and weaknesses. Integrating them into a more general framework for the study of mental disorders is a significant challenge. Symptoms for most disorders in the DSM-IV involve self- or other-report of overt behaviors/emotional states and interpersonal dysfunction, a mixture of which generally amount to clinically significant distress or impairment in some aspect of life. While a section on neurobiological correlates is included for most disorders, none of these is among the actual criteria that lead to diagnosis. The past few years have thus seen a strong push for uncovering biomarkers and molecular genetic components of disorders as well, especially with initiatives such as RDoC and BRAIN which are being heavily promoted by the U.S. National Institutes of Health. The problem here though, is how to combine (or perhaps prioritize) such criteria in relation to overt, self-

report symptoms. An obvious question here is why should one place so much emphasis on integrating results from various methodologies?

One answer is we are all studying related phenomena, whether we hail from psychology, psychiatry, social work, or neuroscience, in part thanks to the common language provided by the DSM. Instead of relying on a single discipline or single methodology, a more appropriate strategy involves identifying concordance between results from different methodologies. Relying on the wrong source, or on only one source, can be misleading. Consider, for example, intelligence in the general population. Intelligence, as measured by IQ tests, is essentially continuously distributed in the general population. Despite apparent continuity and polygenic background (Davies et al., 2011), it is known that low IQ can have discrete causes such as phenylketonuria, Down syndrome, Prader-Willi syndrome, and brain damage, to name but a few. It is very easy to see here how prioritizing just one methodology – overt measurement – in understanding intelligence is misleading. On the one hand, this may seem like an obvious recommendation. But on the other, what occurs often in psychology in that results from one domain tend to lead those in another. For example, given the lack of neurobiological correlates for mental disorders, most research in this field has generally been unidirectional, from disorder or personality constructs to their underlying neurobiology or genetics, with an implicit assumption that this is the right way to go (e.g., genes for X disorder, or brain activation in relation to some personality trait). The problem is not solved merely by conducting research in the opposite direction – e.g., tracing genes and brain circuits, and buliding up larger constructs from them. Rather, what is required, is a true synthesis of information – a strategy that looks for parallels in findings from multiple domains. That is to say, hypotheses generated in one must be tested in the other. For example, if a disorder or personality trait is considered to be continuous (as several proposals for various DSM domains have suggested), then one must be able to find a similarly continuous biological process that corresponds to this trait; or if not, the way in which such a discontinuous or perhaps nonlinear process translates into a continuous trait. We have attempted to provide examples of this strategy in the two disorder sections provided below.

While there are scientific reasons for integrating across disciplines and methodologies, there are also more pragmatic ones. For example, given the somewhat divergent paths of the RDoC and the DSM-5, if research on disorders as delineated by the DSM-5 are not prioritized for federal funding, what implications would this have for future versions of the diagnostic manual? Additionally, while the RDoC appears to have developed its own constructs and strategies for understanding mental disorders, it has done so agnostic of the DSM (NIMH, 2011). Developing a research classification system that is not fully aligned with the DSM has the potential to create confusion in the field, and over time, a schism between researchers and clinicians that might slow the translation of translational research; for this reason, it is important to maintain crosswalks between the two systems. Ultimately, the goal would be to reach a multidisciplinary and multi-perspective understanding of mental disorders, and in this regard the NIMH has announced that its long-term goal with RDoC is to create a research literature that supports future revisions to current diagnostic systems (DSM, ICD) rather than creating a new and competing nosology.

Statistical models and technological advances will be necessary to advance the study of mental disorders, but appropriately harnessing these technologies will require users to have intimate knowledge of the advantages and pitfalls of each methodology, and what conclusions can legitimately be drawn from them. Understandably, it is not possible to list every caveat each time one writes the discussion section of an article, but far more caution is warranted in such instances, and authors (and editors!) should feel less compelled to convince their reader (or reviewer) that their study provides a positive or useful result.

In the following section, we will synthesize information from a variety of literatures to draw broader inferences about the etiology of depression and substance use, and simultaneously address in each example the conceptual issues above.

1. We start with depression, where the results of numerous statistical analyses on self-reported symptoms conclude that depression is the product of a single continuously distributed liability, but evidence from other domains is not entirely consistent with this conjecture. Depression provides an excellent example of how theory-driven analyses and integration of data across multiple sources provides much more information than any single source.
2. We follow with a discussion of the etiology of substance use disorders from the perspective of the gateway and disinhibitory theories. This section focuses on the advantage of causally informative research design and a guiding theory in making meaningful scientific progress.

The main thread connecting these examples is the idea that using a mix of theory and methodology that allows for risky tests is crucial to develop an etiological understanding of these disorders, and that scientific questions cannot be easily or satisfactorily answered using just statistics, or data from one methodology, or even just theory.

3. Understanding Mental Disorders Using Theory and Risky Tests

3.1 Etiology of Major Depression

Major depressive disorder is among the most commonly diagnosed mental disorders with significant personal and societal impact (Bromet et al., 2011; Kessler, 2012). Here we have selectively reviewed the depression literature, unwedded to a preconceived etiological theory, to illustrate how it is possible to transition from an exploratory and inductive knowledge-development phase to a more deductive theory-development phase. This was accomplished by identifying common themes and bridging different research domains to derive a plausible theory regarding the nature of depression that fits the array of findings. The resulting formulation can be further evaluated and refined in subsequent investigations using research designs that provide for risky tests.

Much debate exists on whether depression should be considered a dimensional or categorical condition (Coyne, 1994; Flett, Vredenburg, & Krames, 1997; Prisciandaro & Roberts, 2009) and various subtypes of depression have been proposed over the years (endogenous vs exogenous, bipolar vs unipolar, melancholic vs nonmelancholic, etc.). Given the wealth of literature in this realm, it is not feasible for any one paper to summarize results

of all prior studies. Broadly, however, prominent studies and major reviews (e.g., Andrews et al., 2007; Flett, et al., 1997; Haslam, 2003; Haslam, Holland, & Kuppens, 2012; Kendler & Gardner, 1998; Solomon, Haaga, & Arnow, 2001) undertaken in this area suggest that depression is more likely dimensional in nature, though there does occasionally appear to be evidence for taxonicity (Flett, et al., 1997; Haslam, 2003). Other researchers, however, have argued a case for at least one subtype of depression. For example, Parker and colleagues (2010) have proposed that melancholia (which they define as being similar to endogenous depression) should be classified as a distinct disorder in the upcoming version of the DSM based on various lines of evidence including enhanced responsiveness to biological interventions such as ECT. Others have proposed a two dimensional system – one indexing chronicity, and the other severity (D. N. Klein, 2008). Nevertheless, despite these contentions, in the absence of clear-cut evidence for distinct categories, depression, for now, is still considered by many to be dimensional in nature (Haslam, et al., 2012; Kessler, 2002).

Prior work attempting to study the nature of depression has focused primarily on self-report or interview-based measures of subjective symptomatology (Flett, et al., 1997; Haslam, et al., 2012). However, some laboratory-based measures produce a different picture. Studies that use startle blink reflex and event-related potential (ERP) measures, in conjunction with data from family studies, suggest there may a distinct subtype of depression or at least, a non-linearity at the extreme end of the depression continuum, characterized typically by recurrence. Whether this is equivalent to melancholia or not is not entirely clear based on the information available. The two neurobiological responses – the startle blink reflex response and the error-related negativity (ERN; a brain potential that is thought to reflect error monitoring) – that we discuss in relation to depression were chosen since there is (a) a strong theoretical and experimental literature that document the conditions that give rise to them, (b) the neural circuitry underlying them has been well studied, and (c) they have been investigated in relation to depression.

Among psychophysiological indices, one of the most widely used measures in the study of emotion is the startle blink reflex. Prior work has shown that the modulation of the amplitude of the blink response is directly mediated by the amygdala, and that it is affected by internalizing states such as fear and anxiety (Davis, Walker, Miles, & Grillon, 2009). It has been frequently used to study emotional processing and reactivity in mood and anxiety disorders. Subjects in the general population startle greatest in the context of unpleasant stimuli (such as pictures of mutilated bodies, threatening animals, etc.), and least in the context of pleasant stimuli (images of cute animals and babies, or erotic pictures), both relative to neutral scenes (e.g., household objects and neutral faces), resulting in a linear pattern of response amplitudes (i.e., pleasant < neutral < unpleasant; cf. Vrana, Spence, & Lang, 1988). This increase in response in the context of unpleasant stimuli is referred to as the fear-potentiated startle (FPS) effect.

Several of these kinds of studies have found that patterns of startle blink reactivity in depressed individuals appear to differ from those in the general population. In studies that present startle probes in the context of pictures, those with depression do not show the expected linear pattern of responses. Instead they show more of a flattened pattern of responses with not much differentiation between pleasant and unpleasant responses, or even

increased responses to pleasant pictures. Interestingly, this effect is often restricted to those who have experienced multiple episodes of depression (Forbes, Miller, Cohn, Fox, & Kovacs, 2005), are severely depressed (N. B. Allen, Trinder, & Brennan, 1999), or have high levels of anhedonia (Kaviani et al., 2004). Similarly, in experiments by Lang and colleagues, where subjects are startled in the context of imagined scenes rather than pictures, depressed individuals show a smaller FPS than non-depressed subjects (McTeague et al., 2010; McTeague et al., 2009; McTeague, Lang, Wangelin, Laplante, & Bradley, 2012). Those with a greater number of recurrent episodes also have a smaller FPS than those with just single episodes of depression or those without depression. In our lab (Vaidyanathan, Welo, Malone, Burwell, & Iacono, 2014), we tested this hypothesis directly by contrasting startle response patterns amongst those who had 2 or more episodes of lifetime depression, those with just 1 episode of depression, and those who were never depressed. As hypothesized, only subjects with multiple episodes of depression showed abnormal startle responses. Further, the greater the number of episodes of depression, the stronger the abnormal pattern of responses. Clearly, if a single neurobiological mechanism undergirds all cases of depression, this is not the pattern of results that one would expect.

While most startle studies have tended to focus on current depression, Forbes et al.'s (2005) study revealed that it was specifically lifetime diagnoses that were associated with a lack of startle modulation, rather than current depression. More interestingly, Dichter et al. (2004) found that though individuals with major depression reported improvements in depressive symptoms after twelve weeks of using an antidepressant (Bupropion), their pattern of flat startle reactivity remained the same both before and after taking it. These results suggest that this abnormal startle pattern might be more trait- than state-like. In this context, it is worth noting that meta-analyses (Barbui, Cipriani, Patel, Ayuso-Mateos, & van Ommeren, 2011; Fournier et al., 2010) have suggested that antidepressants are effective relative to placebo only for those who are severely depressed rather than mildly or moderately depressed. In sum, startle studies suggest that there is something qualitatively different about recurrent or chronic depression that leads to decreased emotion modulated startle.

Another psychophysiological index – the error-related negativity (ERN) – also suggests a similar conclusion. The ERN has been used to study the brain's ability to detect errors or monitor one's performance. It is typically observed as a negative deflection in an ERP soon after an individual makes an error in a speeded reaction-time task. Several researchers have posited that the ERN arises from the anterior cingulate cortex (ACC; Dehaene, Posner, & Tucker, 1994; Holroyd, Dien, & Coles, 1998) thought to be involved in self-regulation (E. K. Miller & Cohen, 2001). A variety of studies have examined the ERN in relation to both depression and anxiety, with a greater or more negative ERN evident in anxious individuals (Gehring, Himle, & Nisenson, 2000; Hajcak, McDonald, & Simons, 2003, 2004; Hajcak & Simons, 2002; Johannes et al., 2001; Luu, Collins, & Tucker, 2000; Ruchsow et al., 2005; Stern et al., 2010; Anna Weinberg, Olvet, & Hajcak, 2010). Anxiety and depression are highly comorbid, with many researchers seeing them as distinct expressions of an internalizing spectrum (Krueger, 1999) related to negative affect (Clark & Watson, 1991). Adopting this continuum model, one might expect the results for anxiety disorder to hold for major depression. Contrary to such expectations, however, findings regarding the ERN in major depression have been quite mixed.

While some studies find that individuals with depression appear to show, as with anxiety, a larger ERN (Chiu & Deldin, 2007; Holmes & Pizzagalli, 2008, 2010) than controls, others have reported that depressed patients and controls do not appear to differ significantly in ERN amplitude (Compton et al., 2008; Olvet, Klein, & Hajcak, 2010; Ruchsow et al., 2004; Ruchsow et al., 2006; Schrijvers et al., 2008; Schrijvers et al., 2009). However, Schrijvers et al (2008; 2009) noted that the subjects used in their studies (where no differences were evident between depressed individuals and controls) were more severely depressed than the ones in studies where significant differences were found. Likewise, factors such as anhedonia (Olvet, et al., 2010) and psychomotor retardation (Schrijvers, et al., 2008) (which could be considered indicators of severity) were associated with a reduced or smaller ERN in such studies as well, making subjects with higher levels of these imparing characteristics more comparable to controls, relative to their less severe counterparts who, paradoxically, had larger ERNs. Along similar lines, more Weinberg and colleagues (A. Weinberg, Klein, & Hajcak, 2012) reported that while subjects with generalized anxiety disorder (GAD) alone showed an increased ERN, those with comorbid GAD and depression did not compared to controls; one possibility is that the observed comorbidity is a proxy for the recurrent subtype of depression in their studies. As a whole then, these results suggest that there is a subtype of depression that does not appear to affect the amplitude of the ERN – again, not what would be expected if depression were truly a continuous construct that mapped smoothly on to neurobiology.

Moving on to other domains, one can again see this theme of recurrence carrying over. For example, unlike non-recurrent depression, recurrence appears to occur at equal rates in men and women (Coryell, Endicott, & Keller, 1991; Kessler, McGonagle, Swartz, Blazer, & Nelson, 1993; Kovacs, Obrosky, & Sherrill, 2003; Simpson, Nee, & Endicott, 1997). Dysthymia also appears to be frequently comorbid with recurrent depression (Barkow et al., 2003; Warner, Weissman, Fendrich, Wickramaratne, & Moreau, 1992). Moreover, Klein et al. (2011) in their review on personality traits and depression noted that while negative affect was common to both depression and dysthymia, low positive affect was more specifically associated with dysthymia. They also reported that first-degree relatives of patients with chronic forms of major depression had greater levels of depressive personality traits. Examining the relationship between the course of depression and personality from age 17 to 29, Wilson et al. (2014) found that low positive affect at age 17 predicted the subsequent development of depression, but only for those who went on to have a recurrent course. Burcusa and Iacono (2007) and other researchers (Wichers, Geschwind, van Os, & Peeters, 2010) have observed that though recurrence in depression is a common phenomenon, exactly what causes recurrence is unclear. Additionally, they note that though it is posited that experiencing an episode somehow leaves a scar such that it increases the likelihood of future episodes (i.e., sensitizes or causes a “kindling” effect), there is not much evidence for scarring in prior studies. Indeed, Monroe and Harkness (2005) and Stroud, Davila and Moyer (2008) have found that the first episode of depression is more likely to be preceded by stressful life events rather than recurrent episodes. While this was interpreted as supporting kindling effects or the scar hypothesis, one can also see how the same results could be used to bolster the idea that recurrence breeds true relative to single episodes. Sullivan, Neale, and Kendler (2000) found precisely this in a meta-analysis of five family,

adoption, and twin studies where recurrence was the clearest predictor of familial aggregation of the disorder in all the studies they reviewed as compared to other criteria such as number of symptoms, comorbid disorders, early age of onset, and duration of episodes.

This body of results suggests that recurrent depression may be qualitatively different from single episode depression, perhaps with a different genetic etiology, though our prelimiary efforts in this arena suggested that this might not be the case (Wilson, Vaidyanathan, Miller, McGue, & Iacono, 2014). What risky test could we undertake to evaluate further this possibility? One example would be to examine indices such as the blink reflex and ERN in a prospective study of adolescent identical twins where one twin has depression (the rationale behind such a twin study is detailed in the next section). If the affected twin has abnormal startle or a normal ERN, we would predict that the unaffected cotwin would show the same psychophysiological profile. As the twins mature, we would expect the affected twin to develop recurrent depression. Should the cotwin develop depression, we would expect it to become recurrent as well. In contrast, if recurrent depression reflected some sort of scarring effect, then we would expect the deviant psychophysiology to appear only after recurrent depression was established, and we would expect the two members of the pair to match on deviant psychophysiology only after a point where both had developed recurrent depression.

In summary, though prior work using self-report and diagnostic data (Haslam, et al., 2012; Kendler & Gardner, 1998) favors the conclusion that depression may be a continuous, unitary disorder, consideration of findings from other domains suggest that recurrent depression may not lie on this continuum. Recurrence, as specified in the current DSM, is not included among the criteria used to diagnose a major depressive episode; it can only be used as a specifier. This is an important point, as much of the influential statistical modeling in support of continuum conceptualizations uses cross-sectional epidemiological datasets and life-time or last-year snapshots of symptomatology or diagnosis (e.g., Krueger, 1999; Slade & Watson, 2006; Vollebergh et al., 2001). Developmental context, including onset and recurrence, is ignored in such designs.

To summarize then, while self-report symptoms and statistical models provided a common framework in which to define, organize, and understand depression, integrating information from other domains helps us refine our theoretical notions of the disorder much better. In other words, conceptual issue #1 that we raised earlier comes into obvious play here – i.e., statistical fit and parsimony of models of self-report symptoms can only take a researcher so far. As mentioned earlier, note also that genetic studies of the serotonin transporter gene (Karg, et al., 2011; Munafo, et al., 2009; Risch, 2009) have not shed much light on the etiology of depression, though arguably such studies are more technologically advanced than a simple blink reflex or ERP component (i.e., conceptual issue #3). Theoretical notions of the disorder (e.g., melancholia as a biologically distinct category – conceptual issue #2), along with stronger tests of that theory by integrating information from other domains (i.e., conceptual issue #4) add potentially important nosologically-relevant information not obtainable from statistical analyses of self-report symptoms or diagnoses alone.

3.2 Development of Adolescent Alcohol, Nicotine, and Marijuana Use and Abuse

There are a number of theories on the development and maintenance of substance use disorders covering a variety of perspectives. For the sake of simplicity, we consider a few that provide a concrete example of the main point of this paper – how to leverage the use of different research designs and methods to evaluate evidence for and against theories of mental disorders. Narrowing this focus allows demonstration of how one can derive a statistical model from a scientific theory and, through a combination of proper research design *and* statistical analysis, obtain a *risky* test of the scientific theory.

Nicotine dependence, alcohol dependence, marijuana dependence, and other drug dependence disorders, all correlate moderately (e.g., > .50), leading to the valid conclusion that these disorders share portions of an etiological framework (Krueger et al., 2002). Furthermore, researchers using twin studies have found that these correlations are driven primarily by genetic components (Kendler, Jacobson, Prescott, & Neale, 2003; Kendler, Prescott, Myers, & Neale, 2003; Krueger, et al., 2002; Young, Stallings, Corley, Krauter, & Hewitt, 2000), although the correlational framework and the relative importance of genes and environment significantly changes from adolescence into adulthood (Vrieze, Hicks, Iacono, & McGue, 2012). While an improvement over correlational findings in unrelated individuals, standard twin-based analysis (e.g., models of heritability) provides little insight into the nature of genetic etiology and, improperly used or interpreted, can easily mis-estimate the relative contributions of genes and environment to substance use behaviors (Keller, Coventry, Heath, & Martin, 2005; Plomin, DeFries, & Loehlin, 1977; Scarr & McCartney, 1983).

That individuals who use one drug tend to use another is now well established. Adding to this the fact that adolescents tend to experiment with drugs in temporal order (nicotine -> alcohol -> marijuana -> harder drugs) has led to a prominent theory of adolescent substance use development, termed the “gateway theory” (Kandel, 1975; Kandel & Jessor, 2002). This theory holds that when someone tries a drug, like tobacco or alcohol, the experienced high prompts them to try more drugs, perhaps those that provide stronger highs (Vanyukov et al., 2012). An excellent theory, the question is how one might test this as a causal hypothesis? The fact that adolescents tend to use drugs in a temporal ordering is consistent with the causal hypothesis, but this observation does not provide a risky test. One could think of other reasons for the temporal ordering; for example, the order of use is the same as the order of how difficult it is for an adolescent to obtain the drug and use it without being caught. The gateway theory would also appear to hold that nicotine, alcohol, and marijuana would correlate in adults, but the fact that they do correlate is also not strong evidence for the hypothesis. If the gateway theory survives these tests (temporal ordering; adult correlations) then it is corroborated, but only weakly because these are not risky tests. That is, the confirmatory observations could easily be the result of different causal processes that the tests do not rule out.

The ideal test would be to conduct a randomized controlled trial where one group of randomly assigned adolescents was exposed to a gateway drug (like marijuana) and another group was not. Then, during follow-up, determine the extent to which adolescents from each

group are using other, harder, drugs. Clearly unethical, but there are ways to approximate this experimental state of affairs using natural experiments (Rutter, 2007), and create riskier tests for our theories, including the gateway theory. For example, Keyes and colleagues (2008) tested the association between parental smoking and offspring smoking in a sample of families with adopted and biological offspring. Since adopted offspring are genetically unrelated to the parents, this study can be construed as a reasonable test of the direct environmental impact of parental smoking, even though difficulties remain (e.g., parental smoking is confounded with other parental behaviors). Results indicated that adoptive offspring raised by smokers were somewhat more likely to smoke, but were no more likely to use any other drugs, than adoptive offspring raised by non-smokers – clearly not what we would expect if the gateway hypothesis were true and smokers tend to progress to other drugs. What is more, unlike the adoptive offspring of smokers, the biological offspring of smokers were much more likely to use tobacco, alcohol, and other drugs than biological offspring of non-smokers, suggesting a general genetic vulnerability to substance use and abuse that is largely independent of smoking experiences.

Along the same lines, Irons, McGue, Iacono, and Oetting (2007) used a quasi-experimental genetic method called Mendelian randomization (Ebrahim & Smith, 2008; Smith, 2011) in a longitudinal sample of South Korean adoptees to test the gateway hypothesis of whether environmental exposure to alcohol causes individuals to use other drugs later in life. Some individuals of Asian descent carry deficient variants in the *ALDH2* gene, which metabolizes acetaldehyde to acetic acid. Those with the deficient gene do not metabolize acetaldehyde, a byproduct of alcohol. As a result, acetaldehyde builds up as a toxin in the body and causes a variety of unpleasant reactions such as tachycardia, flushing, and nausea. The result is that having the deficient version of *ALDH2* is very protective for alcoholism (odds ratios confidence intervals of .06 to .30; Luczak, Glatt, & Wall, 2006), and those that have the deficient version of the gene will be exposed to less alcohol (because they refrain from drinking it) than those with the functional version. Thus, variation in this gene is analogous to an independent variable in an experiment—it is a proxy for environmental exposure to alcohol that is essentially randomized, in that it is unrelated to other environmental or biological risk factors for alcohol or drug use, especially in a sample where the parents are Caucasian (i.e., no one has the deficient gene) and biologically unrelated (it's an adoption study). Irons et al. (2007) found that those individuals protected from alcohol exposure (effect size for a drinking index was .40), by virtue of having a deficient *ALDH2* gene, were no more or less likely to use other drugs. The result is once again inconsistent with the gateway hypothesis. Other work has found analogous results for nicotine exposure, although using different genes as proxies for nicotine exposure (Vrieze, McGue, & Iacono, 2012). From these few studies alone, one can see that the natural experiments allowed more direct testing of the gateway hypothesis. The gateway hypothesis makes a prediction about an environmental cause of substance use. Both the adoption study and the Mendelian randomization study allow a quasi-experimental control of environmental versus genetic influences, and for this reason the tests are far riskier than those that depend on correlational data found in cross-sectional or longitudinal designs.

Of course, the gateway hypothesis is not falsified by such studies; it can be modified to be consistent with these findings (i.e., theorists could develop ad hoc hypotheses to deal with negative findings – within reasonable limits, of course; Meehl, 1990). For example, perhaps if you get drunk just once, this may be sufficient to potentiate the gateway effect. In that case, using *ALDH2* for a Mendelian randomization study will not work, as deficiencies in *ALDH2* do not prevent individuals from getting drunk once, they just discourage them from using as much and as often as others. In regards to the Keyes et al. (2008) adoption study of parental smoking, maybe tobacco is not a gateway drug, in part because it provides no obviously euphoric high. Future research will have to expand on these findings for alcohol and tobacco to marijuana and other drugs. To date, however, the gateway hypothesis passes only the relatively equivocal tests and fails riskier tests, those that approximate experimental designs (although work continues; Levine et al., 2011). The take-home point is that statistical modeling of correlational, non-experimental data is still correlational—it does not answer questions of causality, regardless of how many paths are drawn, models are fit, or fit indices are satisfied. Convincing nosological systems are those that are based on corroborated etiological theories, preferably with identified causal agents such as genes, pathogens, or environmental stressors that are developmentally informed. Obtaining this knowledge very often requires (quasi-) experimental research, and valid measures of the phenotype, genotype, and environment. Technological advances can contribute greatly to the measurement issue, but do not by themselves provide causal insight without appropriate research design and experimental controls.

The gateway theory is far from falsified at this point and clearly more work is required, hopefully using increasingly savvy designs and risky tests. In fact, one of the great virtues of the gateway theory is that it makes causal predictions that can be rigorously tested. However, let us assume for the sake of discussion that it has been falsified; what alternative theory could integrate and explain existing findings about the development of adolescent substance use? A plausible alternative could be termed a ‘common liability’ model, which posits that a general vulnerability to substance experimentation/use/addiction is an important component of adolescent use development (Vanyukov, et al., 2012). Our group, and others, theorize that the common liability to addiction can be substantially ascribed to individual differences in adolescent behavioral disinhibition (Iacono, Malone, & McGue, 2008; Zucker, Heitzeg, & Nigg, 2011), a broad concept that includes impulse control, incentive salience, and executive function (like planfulness). Deficits in these areas are posited to put youths at risk for experimenting, using, and eventually forming addiction to, substances like nicotine, alcohol, marijuana and other drugs. The theory posits that disinhibition is a neurodevelopmental trait with a substantial genetic etiology, and that children with greater deficits are at higher risk for use of substances generally, as well as for impulsive and antisocial behavior. The manifestation is heterotypic, in that children with these same disinhibitory deficits are at greater risk for oppositional behavior, ADHD traits, and difficulties in behavioral control. As the children age, these genetically driven deficits translate into increased risk for substance use experimentation and eventual addiction. While the theory places significant emphasis on the common liability (behavioral disinhibition), it also posits important drug-specific liabilities such as drug reward pathway susceptibility, peer influences, or drug availability. Indeed, it appears that the role of common liabilities is

more important in adolescence, though drug-specific liabilities become more critical in early adulthood, at least for common substances (Vrieze, McGue, Miller, & Iacono, 2012). However, the point here is simply that a good theory provides the framework from which risky tests can be devised and experiments conducted. Without the theory we are blindly mining data.

Discussion of the common liability model leads to an illustrative conundrum in addiction research: why do individuals with histories of substance use and dependence suffer from poorer cognitive, interpersonal, and occupational dysfunction, as well as differences in neuroanatomy and function, compared to those without substance use histories? One possibility is that this result can be explained by the common liability theory, which would hold that poorer outcomes in adulthood are actually a heterotypic function of premorbid deficits. That is, if you have the common-liability-related deficits in executive function and behavioral control in adolescence, those deficits will continue into adulthood, whether or not you abuse psychoactive substances. Substance use may very well exacerbate those deficits, but the degree to which it does is an open question. Another theory, which might be called a toxicity theory, predicts that the role of premorbid deficits is small, but that negative outcomes for individuals with substance dependence histories is a result of brain damage from the substances. Which theory finds support from the evidence is far from a settled issue, the discussion of which provides us an opportunity to illustrate our conceptual issues discussed above, especially the use of theory and technological advances such as neuroimaging to address causal questions (i.e., the cause of negative outcomes for individuals with substance dependence histories).

Neuroimaging has shown that administration of nearly all abused psychoactive substances activate dopaminergic pathways in the brain (Parvaz, Alia-Klein, Woicik, Volkow, & Goldstein, 2011) including the prefrontal cortex (Goldstein & Volkow, 2002). These findings are the result of controlled experimental designs involving randomized controlled trials of substance administration followed by neuroimaging, as well as studies of individuals who are administered a substance in the midst of a brain scan (e.g., Volkow et al., 1988). A vast amount of neuroimaging research has also demonstrated, through contrasted groups designs, that the brains of those who become addicted to substances (like alcohol or cocaine) are in many ways different from the brains of those who do not become addicted to substances (Goldstein & Volkow, 2002). These results might seem to favor the toxicity hypothesis because the contrasted groups are selected based on their use of a substance, but this is far from clear. A contrasted groups design might show, beyond any doubt, that the brains of cocaine addicts are different from non-addicts, but they provide little to no evidence about *why* the brains are different. It may be that cocaine addicts have relative deficits in executive function and behavioral control prior to substance use, and these deficits persist or worsen into adulthood. The real utility of the contrasted-groups findings has been to cause scientists to speculate about why the contrasted-groups differences exist. Speculation of this nature has been extremely fruitful in addiction research, as neuroimaging results have driven scientists to create a wealth of etiological scientific theories of drug addiction (e.g., Goldstein & Volkow, 2002; Kalivas & Volkow, 2005; Robinson & Berridge, 1993) involving predictions about specific brain regions and

function. Experimental psychology like this is, for practical and ethical reasons, primarily carried out in infrahuman species such as mice, and results are extrapolated to humans (e.g., Robinson, Gorny, Mitton, & Kolb, 2001; Spear, 2000). What is not clear at present is if and to what extent animal results extend to humans. Similar neuroimaging studies of adolescent humans with alcohol use disorders versus controls have found, for example, decreased hippocampal volumes (De Bellis et al., 2000; Medina, Schweinsburg, Cohen-Zion, Nagel, & Tapert, 2007; Nagel, Barlett, Schweinsburg, & Tapert, 2005; Nagel, Schweinsburg, Phan, & Tapert, 2005), altered parietal cortex and cerebellar responses during working memory tasks, and decreased white matter density (McQueeny et al., 2009), suggesting the possibility of decreased memory function caused by alcohol exposure. The literature on brain differences in adults with alcohol use disorders is even more extensive and suggestive of a causal role of alcohol exposure (Buhler & Mann, 2011). The major limitation of this research exists in all contrasted-groups designs: They are not risky tests. The exposure to alcohol is confounded with all the correlates of alcohol exposure that exist in the general population. It is known from long-term longitudinal studies that those adolescents who go on to use and abuse alcohol express higher levels of pre-morbid antisocial behavior (Hayatbakhsh et al., 2008; King, Iacono, & McGue, 2004; Sartor, Lynskey, Heath, Jacob, & True, 2007), impulsivity (Caspi, Moffitt, Newman, & Silva, 1996), cognitive deficits (Moffitt, 1993; Molina & Pelham, 2003), adversity in their home (Najman et al., 1997) and, most important here, premorbid differences in brain function (Iacono, Carlson, Malone, & McGue, 2002a; Iacono & McGue, 2006). That is, the existence of differences between alcohol-exposed brains and alcohol-naive brains does not provide a risky test of the hypothesis that alcohol exposure causes the observed brain differences, regardless of how technologically sophisticated the brain scan is.

What is required to test the toxicity theory is a (quasi-) experimental test in humans? A good start is with the theory (alcohol toxicity) and the prediction we derive from it. If we predict that adolescent alcohol exposure causes a different adult brain perhaps the most straightforward and obvious design is a prospective, longitudinal, co-twin control study, where monozygotic (MZ) twins are assessed prior to onset of alcohol use (e.g., prior to age 11), during the period of alcohol use (e.g., during high school), and at some point in early adulthood (e.g., after age 23). By nature, each MZ twin provides an experimental control for the other, in that they have (nearly) identical genotypes, as well as shared environmental experiences including rearing environment, school, neighborhood, access to alcohol, etc. If alcohol neurotoxicity is responsible for deleterious outcomes, MZ twins discordant for alcohol exposure should presumably also show discordances in brain structure and function as well as differences in psychosocial function and outcome (IQ, income, occupational success, personality, psychopathology). Alternatively, if premorbid risk factors are responsible for these outcomes, the exposed and unexposed twins should resemble each other on these outcomes, including the brain scans, which may provide the mosts sensitive test of whether and how the twins' brains were different.

While we use alcohol toxicity to provide a concrete example, note that our discussion is applicable to any theory about the environmental causes of psychiatric disorder. Further, while a discordant MZ twin design might be preferred, we can get rough approximations to this quasi-experimental ideal with ordinary siblings. While not everyone has an MZ twin,

most people have siblings or other first-degree relatives, and this practical observation has been used to great advantage in the study of rare neuropsychiatric disorders like schizophrenia. Such family-based controls are beginning to be applied to test the toxicity theory for various substances, using neuroimaging no less (Ersche et al., 2012). The Ersche et al. (2012) study found no striking differences between stimulant abusing probands and their healthy siblings, but both groups differed from non-psychiatric controls, indicating that the observed brain anomalies in the stimulant addicts likely preceded their abuse. Studies like this provide a great start to more rigorously evaluating the toxicity hypothesis.

The guiding theories here are central. If the alcohol toxicity hypothesis fails to explain the observed data, then it is discorroborated (or falsified), and this would have enormous public health and policy implications. If alcohol-use-discordant MZ twins are roughly equal in their outcomes, then it would suggest that premorbid problems are responsible for a significant amount of the morbidity and mortality associated with alcohol use. Alcohol use by itself would be considered less harmful, and intervention would focus less on the alcohol use itself and more on the disinhibition, undercontrol, and executive deficits posited by the common liability model. Preventative services instituted to high-risk children and adolescents would receive higher priority than at present and, by targeting common liability processes, would reduce their risk of developing addiction to many drugs, not just alcohol.

In fact, research addressing questions such as these using an alternative methodology, in this case borrowing again from psychophysiology, already exists. In the substance use literature, an early report that examined the pre-adolescent children of alcoholic fathers marshaled evidence that reduction in P3 amplitude was a candidate endophenotype reflecting genetic risk for alcoholism (Begleiter, Porjesz, Bihari, & Kissin, 1984). P3 waves arise in the ERP when applying laboratory tasks that require attention to and recognition of rare stimuli embedded in a series of more common stimuli. We do not argue the endophenotype candidacy of the P3 here (Euser et al., 2012; Iacono & Malone, 2011; G. A. Miller & Rockstroh, 2013), but only note that it is a technologically refined measure of psychophysiological processes related to information processing efficiency that has been profitably employed to study alcoholism as well as a wide range of psychiatric conditions.

Endophenotypes, such as the P3, are typically promoted for their supposed assistance in genetic association studies, but they have yet to demonstrate their promise for gene finding (Malone, Vaidyanathan, et al., 2014; G. A. Miller & Rockstroh, 2013; Vrieze, Malone, Pankratz, et al., 2014; Vrieze, Malone, Vaidyanathan, et al., 2014). In fact, we recently completed a comprehensive evaluation of endophenotypes using psychophysiological and molecular genetic data from paradigms examining P3, the startle blink response, antisaccade eye tracking performance, skin conductance orienting and habituation, and resting electronencephalogram (EEG) frequency band power. Employing a sample of approximately 5000 people, we subjected these putative endophenotypes to a risky test (Iacono, Malone, Vaidyanathan, & Vrieze, 2014; Iacono, Vaidyanathan, Vrieze, & Malone, 2014; Malone, Burwell, et al., 2014; Malone, Vaidyanathan, et al., 2014; Vaidyanathan, Isen, et al., 2014; Vaidyanathan, Malone, et al., 2014; Vaidyanathan, Malone, Miller, McGue, & Iacono, 2014; Vrieze, Malone, Pankratz, et al., 2014; Vrieze, Malone, Vaidyanathan, et al., 2014). While most of the indices in our sample were heritable in

conventional biometric twin models, in molecular genetic tests, whether we examined GWAS common SNPs, rare variants via exome-sequencing, or used whole genome sequencing, we found little credible evidence that any of the molecular variants was associated with any of the indices. Even when using relaxed statistical significance thresholds that were justified by specifically examining lists of candidate genes and SNPs from prior studies demonstrating empirical support or theoretical relevance for our candiates, we did not find any that proved statistically significant. Such results indicated that despite the *a priori* promise derived from combining two advanced technologies – psychophysiology and molecular genetics – it was still not possible to uncover meaningful insights about genetic variants of possible importance to psychiatric disorders. To the contrary, our findings suggest that endophenotypes are just as complex as psychiatric disorders, and endophenotype genetics are just as complicated as psychiatric genetics. A far greater advantage of endophenotypes, in our opinion, is that they provide new ways to test and refine theories. Research on P3, in particular, is relevant to the evaluation of the disinhibition, gateway, and neurotoxicity theories considered in this section.

As noted earlier, studies using self-report data in genetically informed designs have concluded that a common genetic liability for behavioral disinhibition shared across substance use, childhood disruptive, and antisocial disorders accounts for their covariance (Bornovalova, Hicks, Iacono, & McGue, 2010; Hicks, Krueger, Iacono, McGue, & Patrick, 2004; Krueger, et al., 2002). Work with the P3 endophenotype reinforces the validity of this conclusion because each of these disorders individually has been shown to be associated with reduced P3 amplitude (Iacono, Carlson, Malone, & McGue, 2002b) as have externalizing problem behaviors such as early use of cigarettes (Yoon, Iacono, Malone, & McGue, 2006) and early sexual experience (Iacono & McGue, 2006). Further evidence that the clinical and endophenotype research are converging on the same etiological interpretation derives from reports showing that the general liability for externalizing psychopathology shares covariance with P3 amplitude (Patrick et al., 2006), and this association reflects shared genetic risk (Hicks et al., 2007). In combination, these findings support the conclusion that there are biological factors that account for why the same individual often engages in undersocialized behavior in adolescence and subsequently develops polydrug dependence. They also argue against the gateway hypothesis because reduced P3 amplitude is manifest prior to the initiation of substance use, predicts the subsequent development of all classes of substance dependence, and is associated with externalizing problem behaviors (e.g., conduct disorder, ADHD) that both precede and forecast the eventual development of addiction (Iacono & Malone, 2011).

Studies of P3 amplitude have also shed light on the alcohol neurotoxicity hypothesis. If brain mechanisms involved in the generation of P3 are damaged by drinking, then twins discordant for drinking should show corresponding discordance in P3 amplitude. However, results from prior work contradict this; within pair analyses comparing twins who abuse alcohol to their cotwins who don’t, have shown that the P3 amplitudes of discordant twins are the same, clearly challenging the neurotoxicity hypothesis (Carlson, Iacono, & McGue, 2002; Carlson, Iacono, & McGue, 2004). Another way to evaluate this hypothesis is to determine whether the high heritability observed for P3 amplitude is moderated by a history

of alcohol consumption, thus indicating that alcohol is an environmental toxin altering the genetic influence on brain maturation. A prospective twin study that examined how alcohol consumption through adolescence moderated heritability of P3 amplitude at age 18 failed to find any effect, again refuting the neurotoxicity hypothesis at least for this one brain measure (Perlman, Johnson, & Iacono, 2009).

P3 studies are not without limitations. P3 does not capture all aspects of brain function, so the fact that P3 is unaffected by alcohol consumption does not argue that alcohol is without long-term deleterious effects on brain development. Moreover, as P3 accounts for only 3-4% of the variance in externalizing, it provides only a small part of the picture regarding the neurogenetic basis of the common liability model. Further, P3 is a complex waveform reflecting the aggregate influence of multiple neural generators. Decomposing it into component aspects has the potential to uncover brain activity more tightly tied to externalizing with more specific underlying neurobiology (Gilmore, Malone, Bernat, & Iacono, 2009; Gilmore, Malone, & Iacono, 2010). Given the success of P3, research on other endophenotypes for externalizing is clearly warranted.

To conclude, work with P3 supports the common liability-disinhibition model over the gateway hypothesis. The convergence of these P3 findings with the conclusions drawn from studies using other methodological approaches reviewed earlier (e.g., adoption and Mendelian randomization designs) is consistent with the common liability model. The jury is still out, but we attempted to draw this conclusion on the basis of research designs that controlled experimental variables and created riskier tests, which in turn gave us leverage to draw some conclusions about causality. In addition, we have discussed the great utility of technological advances in addiction research, but with comments on their limitations. Technological advances are not scientific advances until a scientist designs a study that uses that technology to engineer a test (preferably risky) of an etiological theory.

4. Concluding Remarks

We have argued that advances in understanding mental disorders are being stymied. We have discussed a variety of reasons for this: (a) overreliance on statistics and technology with little theoretical guidance, (b) a lack of integration of information across various methodologies that researchers use to study psychopathology, and (c) underutilization of research designs that provide for risky tests. Our point here was not merely to critique prevailing approaches to scientific investigation, but to also provide a constructive solution to some of these issues. Thus, we examined how to better conceptualize two different disorders – depression and substance use – using information from various domains, while keeping in mind the limitations we had outlined.

The key issue we have attempted to tackle is how to better extract meaning or interpret results we get from studies. We have more technological and statistical power than ever to conduct larger and larger studies, with more data at a greater resolution than has ever been available. Between this and the easy availability of software and methods for analyzing almost any kind of data, we have had an explosion of new journals and published articles. And yet, the progress we have made in understanding mental disorders is relatively modest

given the degree of human suffering and their societal cost. This should be a very clear sign that what we lack is not information, but rather how to deal with it. Information, in and of itself, is inherently meaningless. It's the quality of the information that we have in relation to our question, and how we use it that matters. The crux of the problem appears to be overreliance in the power of statistical software and technology. Otis Dudley Duncan (1984) referred to this tendency as "statisticism" and derisively noted that this was the "….notion that computing is synonymous with doing research, the naïve faith that statistics is a complete or sufficient basis for scientific methodology, the superstition that statistical formulas exist for evaluating such things as the relative merits of different substantive theories or the "importance" of the causes of a "dependent variable"; and the delusion that decomposing the covariations of some arbitrary and haphazardly assembled collection of variables can somehow justify not only a "causal model" but also, praise a mark, a "measurement model."

Given the size and scope of datasets available today, running a factor analysis of personality data for 10,000 people or comparing brain scans or EEG responses between two disorders is easy to accomplish – and will likely yield results that are statistically significant but scientifically ambiguous. It is very easy at that point to retrofit a rationale to the findings, or come up with an explanatory hypothesis, and reify the presence of latent factors or the biological bases of disorders. This approach is reminiscent of what Richard Feynman (1974) referred to as "cargo cult science". While such a procedure can be made to appear scientific superficially, chances of it leading to useful findings, theoretically or practically, are not very high. It is of course possible that we could stumble into some result at some point if we continue to conduct research this way. Alternatively, we could follow a more elegant and efficient route to useful findings if we thought a bit more critically about our constructs, methodologies, study designs, and statistics before running the next study.

Acknowledgments

Preparation of this manuscript was supported by the National Institute of Drug Abuse (NIDA) grants DA 024417, DA 05147, DA 13240, DA 034606, DA 036216, National Institute of Alcohol Abuse and Alcoholism (NIAAA) grants AA 023974, AA 09367, and the National Institute of Mental Health (NIMH) grant MH 66140. We would also like to thank Brian Hicks, Sylia Wilson, Scott Burwell, and Chris Patrick for their comments on earlier versions of this manuscript.

References

Abecasis GR, Auton A, Brooks LD, DePristo MA, Durbin RM, Handsaker RE, McVean GA, et al. An integrated map of genetic variation from 1,092 human genomes. Nature. 2012; 491(7422):56–65.10.1038/nature11632 [PubMed: 23128226]

Abramowitz Y, Perlman G, Levy E, Beeri R, Chajek-Shaul T, Leibowitz D. Ockham's razor or Hickam's dictum: a right atrial mass following excision of left atrial myxoma. European Journal of Echocardiography. 2008; 9(3):415–416.10.1093/ejechocard/jen023 [PubMed: 18490342]

Achenbach, TM. Integrative Guide to the 1991 CBCL/4-18, YSR, and TRF Profiles. Burlington, VT: University of Vermont; 1991.

Allen HL, Estrada K, Lettre G, Berndt SI, Weedon MN, Rivadeneira F, et al. Consortium, P. Hundreds of variants clustered in genomic loci and biological pathways affect human height. Nature. 2010; 467(7317):832–838.10.1038/Nature09410 [PubMed: 20881960]

Allen NB, Trinder J, Brennan C. Affective startle modulation in clinical depression: preliminary findings. Biological Psychiatry. 1999; 46(4):542–550. [PubMed: 10459405]

Andrews G, Brugha T, Thase ME, Duffy FF, Rucci P, Slade T. Dimensionality and the category of major depressive episode. International Journal of Methods in Psychiatric Research. 2007; 16(S1):S41–S51.10.1002/mpr.216 [PubMed: 17623394]

Barbui C, Cipriani A, Patel V, Ayuso-Mateos JL, van Ommeren M. Efficacy of antidepressants and benzodiazepines in minor depression: systematic review and meta-analysis. The British Journal of Psychiatry. 2011; 198(1):11–16.10.1192/bjp.bp.109.076448 [PubMed: 21200071]

Barkow K, Maier W, Ustun TB, Gonsicke M, Wittchen HU, Heun R. Risk factors for depression at 12-month follow-up in adult primary health care patients with major depression: an international prospective study. Journal of Affective Disorders. 2003; 76(1-3):157–169. [PubMed: 12943946]

Begleiter H, Porjesz B, Bihari B, Kissin B. Event-related brain potentials in boys at risk for alcoholism. Science. 1984; 225(4669):1493–1496. [PubMed: 6474187]

Behrens JT. Principles and procedures of exploratory data analysis. Psychological Methods. 1997; 2(2):131–160.10.1037//1082-989x.2.2.131

Bornovalova MA, Hicks BM, Iacono WG, McGue M. Familial transmission and heritability of childhood disruptive disorders. Am J Psychiatry. 2010; 167(9):1066–1074.10.1176/appi.ajp.2010.09091272 [PubMed: 20634367]

Bouchard TJ, Lykken DT, Mcgue M, Segal NL, Tellegen A. Sources of Human Psychological Differences - the Minnesota Study of Twins Reared Apart. Science. 1990; 250(4978):223–228. [PubMed: 2218526]

Breiman L. Bagging predictors. Machine Learning. 1996; 24(2):123–140.

Breiman L. Random forests. Machine Learning. 2001a; 45(1):5–32.

Breiman L. Statistical modeling: The two cultures. Statistical Science. 2001b; 16(3):199–215.

Bromet E, Andrade L, Hwang I, Sampson N, Alonso J, de Girolamo G, Kessler R, et al. Cross-national epidemiology of DSM-IV major depressive episode. BMC Medicine. 2011; 9(1):90. [PubMed: 21791035]

Buhler M, Mann K. Alcohol and the Human Brain: A Systematic Review of Different Neuroimaging Methods. Alcoholism-Clinical and Experimental Research. 2011; 35(10):1771–1793.10.1111/j.1530-0277.2011.01540.x

Burcusa SL, Iacono WG. Risk for recurrence in depression. Clinical Psychology Review. 2007; 27(8):959–985.10.1016/j.cpr.2007.02.005 [PubMed: 17448579]

Button KS, Ioannidis JPA, Mokrysz C, Nosek BA, Flint J, Robinson ESJ, Munafo MR. Power failure: why small sample size undermines the reliability of neuroscience. Nat Rev Neurosci. 2013; 14(5):365–376. [PubMed: 23571845]

Cameron, WB. Informal sociology: A casual introduction to sociological thinking. Vol. 21. Random House; New York: 1963.

Carlson SR, Iacono WG, McGue M. P300 amplitude in adolescent twins discordant and concordant for alcohol use disorders. Biological Psychology. 2002; 61:203–227. [PubMed: 12385676]

Carlson SR, Iacono WG, McGue M. P300 amplitude in non-alcoholic adolescent twin pairs who become discordant for alcoholism as adults. Psychophysiology. 2004; 41:841–844. [PubMed: 15563337]

Caspi A, Moffitt TE, Newman DL, Silva PA. Behavioral observations at age 3 years predict adult psychiatric disorders. Longitudinal evidence from a birth cohort. Arch Gen Psychiatry. 1996; 53(11):1033–1039. [PubMed: 8911226]

Caspi A, Sugden K, Moffitt TE, Taylor A, Craig IW, Harrington H, Poulton R, et al. Influence of life stress on depression: Moderation by a polymorphism in the 5-HTT gene. Science. 2003; 297:851–854. [PubMed: 12161658]

Chiu PH, Deldin PJ. Neural evidence for enhanced error detection in major depressive disorder. The American Journal of Psychiatry. 2007; 164(4):608–616. [PubMed: 17403974]

Claeskens, G.; Hjort, NL. Model Selection and Model Averaging. Cambridge: Cambridge University Press; 2008.

Clark LA, Watson D. Tripartite model of anxiety and depression: Psychometric evidence and taxonomic implications. Journal of Abnormal Psychology. 1991; 100(3):316–336.10.1037/0021-843x.100.3.316 [PubMed: 1918611]

Collins AL, Kim Y, Sklar P, O'Donovan MC, Sullivan PF. Hypothesis-driven candidate genes for schizophrenia compared to genome-wide association results. Psychol Med. 2012; 42(3):607–616.10.1017/S0033291711001607 [PubMed: 21854684]

Compton RJ, Lin M, Vargas G, Carp J, Fineman SL, Quandt LC. Error detection and posterror behavior in depressed undergraduates. Emotion. 2008; 8(1):58–67.10.1037/1528-3542.8.1.58 [PubMed: 18266516]

Coryell W, Endicott J, Keller MB. Predictors of relapse into major depressive disorder in a nonclinical population. The American Journal of Psychiatry. 1991; 148(10):1353–1358. [PubMed: 1897616]

Coyne JC. Self-reported distress: Analog or ersatz depression? Psychological Bulletin. 1994; 116(1): 29–45.10.1037/0033-2909.116.1.29 [PubMed: 8078972]

Cronbach LJ. The two disciplines of scientific psychology. American psychologist. 1957; 12(11):671.

Cronbach LJ, Meehl PE. Construct validity in psychological tests. Psychological Bulletin. 1955; 52(4): 281–302. [PubMed: 13245896]

D'Onofrio BM, Turkheimer EN, Eaves LJ, Corey LA, Berg K, Solaas MH, Emery RE. The role of the Children of Twins design in elucidating causal relations between parent characteristics and child outcomes. Journal of Child Psychology and Psychiatry and Allied Disciplines. 2003; 44(8):1130–1144.

Davies G, Tenesa A, Payton A, Yang J, Harris SE, Liewald D, Deary IJ, et al. Genome-wide association studies establish that human intelligence is highly heritable and polygenic. Mol Psychiatry. 2011; 16(10):996–1005. doi: http://www.nature.com/mp/journal/v16/n10/suppinfo/mp201185s1.html. [PubMed: 21826061]

Davis M, Walker DL, Miles L, Grillon C. Phasic vs sustained fear in rats and humans: Role of the extended amygdala in fear vs anxiety. Neuropsychopharmacology. 2009; 35(1):105–135. [PubMed: 19693004]

De Bellis MD, Clark DB, Beers SR, Soloff PH, Boring AM, Hall J, Keshavan MS, et al. Hippocampal volume in adolescent-onset alcohol use disorders. Am J Psychiatry. 2000; 157(5):737–744. [PubMed: 10784466]

Dehaene S, Posner MI, Tucker DM. Localization of a neural system for error detection and compensation. Psychological Science. 1994; 5(5):303–305.10.1111/j.1467-9280.1994.tb00630.x

Dichter GS, Tomarken AJ, Shelton RC, Sutton SK. Early- and late-onset startle modulation in unipolar depression. Psychophysiology. 2004; 41(3):433–440.10.1111/j.1469-8986.00162.x [PubMed: 15102129]

Digman JM. Personality Structure - Emergence of the 5-Factor Model. Annual Review of Psychology. 1990; 41:417–440.

Duncan LE, Keller MC. A Critical Review of the First 10 Years of Candidate Gene-by-Environment Interaction Research in Psychiatry. American Journal of Psychiatry. 2011; 168(10):1041–1049.10.1176/appi.ajp.2011.11020191 [PubMed: 21890791]

Duncan, OD. Notes on social measurement: Historical and critical. Vol. 1. Russell Sage Foundation Publications; 1984.

Ebrahim S, Smith GD. Mendelian randomization: can genetic epidemiology help redress the failures of observational epidemiology? Human Genetics. 2008; 123(1):15–33.10.1007/s00439-007-0448-6 [PubMed: 18038153]

Ersche KD, Jones PS, Williams GB, Turton AJ, Robbins TW, Bullmore ET. Abnormal Brain Structure Implicated in Stimulant Drug Addiction. Science. 2012; 335(6068):601–604.10.1126/science.1214463 [PubMed: 22301321]

Euser AS, Arends LR, Evans BE, Greaves-Lord K, Huizink AC, Franken IH. The P300 event-related brain potential as a neurobiological endophenotype for substance use disorders: a meta-analytic investigation. Neurosci Biobehav Rev. 2012; 36(1):572–603.10.1016/j.neubiorev.2011.09.002 [PubMed: 21964481]

Farah MJ. Brain Images, Babies, and Bathwater: Critiquing Critiques of Functional Neuroimaging. Hastings Center Report. 2014; 44(s2):S19–S30.10.1002/hast.295 [PubMed: 24634081]

Feynman RP. Cargo cult science. Engineering and Science. 1974; 37(7):10–13.

Flett GL, Vredenburg K, Krames L. The continuity of depression in clinical and nonclinical samples. Psychological Bulletin. 1997; 121(3):395–416.10.1037/0033-2909.121.3.395 [PubMed: 9136642]

Forbes EE, Miller A, Cohn JF, Fox NA, Kovacs M. Affect-modulated startle in adults with childhood-onset depression: Relations to bipolar course and number of lifetime depressive episodes. Psychiatry Research. 2005; 134(1):11–25. [PubMed: 15808286]

Fournier JC, DeRubeis RJ, Hollon SD, Dimidjian S, Amsterdam JD, Shelton RC, Fawcett J. Antidepressant drug effects and depression severity: A patient-level meta-analysis. JAMA: Journal of the American Medical Association. 2010; 303(1):47–53.10.1001/jama.2009.1943 [PubMed: 20051569]

Fu WQ, O'Connor TD, Jun G, Kang HM, Abecasis G, Leal SM, et al. Project, N. E. S. Analysis of 6,515 exomes reveals the recent origin of most human protein-coding variants. Nature. 2013; 493(7431):216–220.10.1038/Nature11690 [PubMed: 23201682]

Galton F. The history of twins, as a criterion of the relative powers of nature and nurture. Fraser's Magazine. 1875; 12:566–576.

Gehring WJ, Himle J, Nisenson LG. Action-Monitoring Dysfunction in Obsessive-Compulsive Disorder. Psychological Science. 2000; 11(1):1–6.10.1111/1467-9280.00206 [PubMed: 11228836]

Gilmore CS, Malone SM, Bernat EM, Iacono WG. Relationship between the P3 event-related potential, its associated time-frequency components, and externalizing psychopathology. Psychophysiology. 200910.1111/j.1469-8986.2009.00876.x

Gilmore CS, Malone SM, Iacono WG. Brain electrophysiological endophenotypes for externalizing psychopathology: a multivariate approach. Behav Genet. 2010; 40(2):186–200.10.1007/s10519-010-9343-3 [PubMed: 20155392]

Goldstein RZ, Volkow ND. Drug addiction and its underlying neurobiological basis: Neuroimaging evidence for the involvement of the frontal cortex. American Journal of Psychiatry. 2002; 159(10):1642–1652. [PubMed: 12359667]

Gonzalez-Castillo J, Saad ZS, Handwerker DA, Inati SJ, Brenowitz N, Bandettini PA. Whole-brain, time-locked activation with simple tasks revealed using massive averaging and model-free analysis. Proceedings of the National Academy of Sciences. 2012; 109(14):5487–5492.10.1073/pnas.1121049109

Grove, WM.; Vrieze, SI. On the substantive grounding and clinical utility of categories versus dimensions. In: Millon, T.; Krueger, RF.; Simonsen, E., editors. Contemporary directions in psychopathology: Scientific foundations of the DSM-V and ICD-11. New York, NY, US: Guilford Press; 2010. p. 303-323.

Hajcak G, McDonald N, Simons RF. Anxiety and error-related brain activity. Biological Psychology. 2003; 64(1–2):77–90. [PubMed: 14602356]

Hajcak G, McDonald N, Simons RF. Error-related psychophysiology and negative affect. Brain and Cognition. 2004; 56(2):189–197. [PubMed: 15518935]

Hajcak G, Simons RF. Error-related brain activity in obsessive-compulsive undergraduates. Psychiatry Research. 2002; 110(1):63–72. [PubMed: 12007594]

Hart AB, de Wit H, Palmer AA. Candidate Gene Studies of a Promising Intermediate Phenotype: Failure to Replicate. Neuropsychopharmacology. 2013; 38(5):802–816. [PubMed: 23303064]

Haslam N. Categorical versus dimensional models of mental disorder: the taxometric evidence. Australian and New Zealand Journal of Psychiatry. 2003; 37(6):696–704.10.1111/j.1440-1614.2003.01258.x [PubMed: 14636384]

Haslam N, Holland E, Kuppens P. Categories versus dimensions in personality and psychopathology: a quantitative review of taxometric research. Psychological Medicine. 2012; 42(05):903–920.10.1017/S0033291711001966 [PubMed: 21939592]

Hayatbakhsh MR, Najman JM, Jamrozik K, Al Mamun A, Bor W, Alati R. Adolescent problem behaviours predicting DSM-IV diagnoses of multiple substance use disorder. Findings of a prospective birth cohort study. Soc Psychiatry Psychiatr Epidemiol. 2008; 43(5):356–363.10.1007/s00127-008-0325-1 [PubMed: 18301851]

Hicks BM, Bernat E, Malone SM, Iacono WG, Patrick CJ, Krueger RF, McGue M. Genes mediate the association between P3 amplitude and externalizing disorders. Psychophysiology. 2007; 44(1):98–105. [PubMed: 17241145]

Hicks BM, Krueger RF, Iacono WG, McGue M, Patrick CJ. Family transmission and heritability of externalizing disorders: A twin-family study. Archives of General Psychiatry. 2004; 61:922–928. [PubMed: 15351771]

Holland PW. Statistics and Causal Inference. Journal of the American Statistical Association. 1986; 81(396):945–960.

Holmes A, Pizzagalli DA. Spatiotemporal dynamics of error processing dysfunctions in major depressive disorder. Archives of General Psychiatry. 2008; 65(2):179–188. [PubMed: 18250256]

Holmes A, Pizzagalli DA. Effects of task-relevant incentives on the electrophysiological correlates of error processing in major depressive disorder. Cognitive, Affective, & Behavioral Neuroscience. 2010; 10(1):119–128.10.3758/cabn.10.1.119

Holroyd CB, Dien J, Coles MGH. Error-related scalp potentials elicited by hand and foot movements: evidence for an output-independent error-processing system in humans. Neuroscience Letters. 1998; 242(2):65–68. [PubMed: 9533395]

Iacono WG, Carlson SR, Malone SM, McGue M. P3 event-related potential amplitude and the risk for disinhibitory disorders in adolescent boys. Archives of General Psychiatry. 2002a; 59(8):750–757. [PubMed: 12150652]

Iacono WG, Carlson SR, Malone SM, McGue M. P3 event-related potential amplitude and the risk for disinhibitory disorders in adolescent boys. Arch Gen Psychiatry. 2002b; 59(8):750–757. [PubMed: 12150652]

Iacono WG, Malone SM. Developmental Endophenotypes: Indexing Genetic Risk for Substance Abuse With the P300 Brain Event-Related Potential. Child Development Perspectives. 2011; 5(4): 239–247.10.1111/j.1750-8606.2011.00205.x [PubMed: 22247735]

Iacono WG, Malone SM, McGue M. Behavioral disinhibition and the development of early-onset addiction: Common and specific influences. Annual Review of Clinical Psychology. 2008; 4:325–348.10.1146/annurev.clinpsy.4.022007.141157

Iacono WG, Malone SM, Vaidyanathan U, Vrieze SI. Genome-wide scans of genetic variants for psychophysiological endophenotypes: A methodological overview. Psychophysiology. 2014; 51(12):1207–1224. [PubMed: 25387703]

Iacono WG, McGue M. Association between P3 event-related brain potential amplitude and adolescent problem behavior. Psychophysiology. 2006; 43:465–469. [PubMed: 16965608]

Iacono WG, Vaidyanathan U, Vrieze SI, Malone SM. Knowns and unknowns for psychophysiological endophenotypes: Integration and response to commentaries. Psychophysiology. 2014; 51(12): 1339–1347. [PubMed: 25387720]

Irons DE, McGue M, Iacono WG, Oetting WS. Mendelian randomization: A novel test of the gateway hypothesis and models of gene-environment interplay. Development and Psychopathology. 2007; 19(4):1181–1195.10.1017/S0954579407000612 [PubMed: 17931442]

Johannes S, Wieringa BM, Nager W, Rada D, Dengler R, Emrich HM, Dietrich DE, et al. Discrepant target detection and action monitoring in obsessive-compulsive disorder. Psychiatry Research: Neuroimaging. 2001; 108(2):101–110. [PubMed: 11738544]

Jöreskog, KG.; Sörbom, D. PRELIS: A Preprocessor for LISREL. Mooresville: Indiana: Scientific Software; 1986.

Kalivas PW, Volkow ND. The neural basis of addiction: A pathology of motivation and choice. American Journal of Psychiatry. 2005; 162(8):1403–1413. [PubMed: 16055761]

Kandel DB. Stages in Adolescent Involvement in Drug-Use. Science. 1975; 190(4217):912–914. [PubMed: 1188374]

Kandel, DB.; Jessor, R. The gateway hypothesis revisited. In: Kandel, DB., editor. Stages and Pathways of Drug Involvement. New York: Cambridge University Press; 2002.

Karg K, Burmeister M, Shedden K, Sen S. The serotonin transporter promoter variant (5-HTTLPR), stress, and depression meta-analysis revisited: evidence of genetic moderation. Arch Gen Psychiatry. 2011; 68(5):444–454.10.1001/archgenpsychiatry.2010.189 [PubMed: 21199959]

Kaviani H, Gray JA, Checkley SA, Raven PW, Wilson GD, Kumari V. Affective modulation of the startle response in depression: influence of the severity of depression, anhedonia, and anxiety. Journal of Affective Disorders. 2004; 83(1):21–31. [PubMed: 15546642]

Keller MC, Coventry WL, Heath AC, Martin NG. Widespread evidence for non-additive genetic variation in Cloninger's and Eysenck's personality dimensions using a twin plus sibling design. Behavior Genetics. 2005; 35(6):707–721.10.1007/s10519-005-6041-7 [PubMed: 16273321]

Kendler KS, Gardner CO. Boundaries of major depression: an evaluation of DSM-IV criteria. American Journal of Psychiatry. 1998; 155(2):172–177. [PubMed: 9464194]

Kendler KS, Jacobson KC, Prescott CA, Neale MC. Specificity of genetic and environmental risk factors for use and abuse/dependence of cannabis, cocaine, hallucinogens, sedatives, stimulants, and opiates in male twins. American Journal of Psychiatry. 2003; 160(4):687–695. [PubMed: 12668357]

Kendler KS, Prescott CA, Myers J, Neale MC. The structure of genetic and environmental risk factors for common psychiatric and substance use disorders in men and women. Archives of General Psychiatry. 2003; 60(9):929–937. [PubMed: 12963675]

Kessler RC. The categorical versus dimensional assessment controversy in the sociology of mental illness. Journal of Health and Social Behavior. 2002:171–188. [PubMed: 12096698]

Kessler RC. The costs of depression. The Psychiatric clinics of North America. 2012; 35(1):1–14. [PubMed: 22370487]

Kessler RC, McGonagle KA, Swartz M, Blazer DG, Nelson CB. Sex and depression in the National Comorbidity Survey. I: Lifetime prevalence, chronicity and recurrence. Journal of Affective Disorders. 1993; 29(2-3):85–96. [PubMed: 8300981]

Keyes MA, Legrand LN, Iacono WG, McGue M. Parental smoking and adolescent problem behavior: An adoption study of general and specific effects. American Journal of Psychiatry. 2008; 165(10): 1338–1344.10.1176/appi.ajp.2008.08010125 [PubMed: 18676589]

Kilpatrick DG, Koenen KC, Ruggiero KJ, Acierno R, Galea S, Resnick HS, Gelernter J, et al. The serotonin transporter genotype and social support and moderation of posttraumatic stress disorder and depression in hurricane-exposed adults. American Journal of Psychiatry. 2007; 164(11):1693–1699.10.1176/appi.ajp.2007.06122007 [PubMed: 17974934]

King SM, Iacono WG, McGue M. Childhood externalizing and internalizing psychopathology in prediction of early substance use. Addiction. 2004; 99:1548–1559. [PubMed: 15585046]

Klein DN. Classification of Depressive Disorders in the DSM-V: Proposal for a Two-Dimension System. Journal of Abnormal Psychology. 2008; 117(3):552–560. [PubMed: 18729608]

Klein DN, Kotov R, Bufferd SJ. Personality and depression: Explanatory models and review of the evidence. Annual Review of Clinical Psychology. 2011; 7(1):269–295.10.1146/annurev-clinpsy-032210-104540

Kovacs M, Obrosky DS, Sherrill J. Developmental changes in the phenomenology of depression in girls compared to boys from childhood onward. Journal of Affective Disorders. 2003; 74(1):33–48. [PubMed: 12646297]

Krueger RF. The structure of common mental disorders. Archives of General Psychiatry. 1999; 56(10): 921–926. [PubMed: 10530634]

Krueger RF, Hicks BM, Patrick CJ, Carlson SR, Iacono WG, McGue M. Etiologic connections among substance dependence, antisocial behavior, and personality: Modeling the externalizing spectrum. Journal of Abnormal Psychology. 2002; 111(3):411–424. [PubMed: 12150417]

Kupfer DJ, Regier DA. Neuroscience, clinical evidence, and the future of psychiatric classification in DSM-5. American Journal of Psychiatry. 2011; 168(7):672–674. [PubMed: 21724672]

Lakatos, I. The Methodology of Scientific Research Programmes. Vol. 1. Cambridge: Cambridge University Press; 1978.

Lander ES. Initial impact of the sequencing of the human genome. Nature. 2011; 470(7333):187–197.10.1038/nature09792 [PubMed: 21307931]

Levine A, Huang Y, Drisaldi B, Griffin EA Jr, Pollak DD, Xu S, Kandel ER, et al. Molecular mechanism for a gateway drug: epigenetic changes initiated by nicotine prime gene expression by cocaine. Sci Transl Med. 2011; 3(107):107ra109.10.1126/scitranslmed.3003062

Logothetis NK. What we can do and what we cannot do with fMRI. Nature. 2008; 453(7197):869–878. http://www.nature.com/nature/journal/v453/n7197/suppinfo/nature06976_S1.html. 10.1038/nature06976 [PubMed: 18548064]

Luczak SE, Glatt SJ, Wall TL. Meta-analyses of ALDH2 and ADH1B with alcohol dependence in Asians. Psychological Bulletin. 2006; 132(4):607–621. [PubMed: 16822169]

Luu P, Collins P, Tucker DM. Mood, personality, and self-monitoring: Negative affect and emotionality in relation to frontal lobe mechanisms of error monitoring. Journal of Experimental Psychology: General. 2000; 129(1):43–60.10.1037/0096-3445.129.1.43 [PubMed: 10756486]

Malone SM, Burwell SJ, Vaidyanathan U, Miller MB, Mcgue M, Iacono WG. Heritability and molecular-genetic basis of resting EEG activity: A genome-wide association study. Psychophysiology. 2014; 51(12):1225–1245. [PubMed: 25387704]

Malone SM, Vaidyanathan U, Basu S, Miller MB, Mcgue M, Iacono WG. Heritability and molecular-genetic basis of the P3 event-related brain potential: A genome-wide association study. Psychophysiology. 2014; 51(12):1246–1258. [PubMed: 25387705]

Maraun MD. Metaphor taken as math: Indeterminacy in the factor analysis model. Multivariate Behavioral Research. 1996; 31(4):517–538. [PubMed: 26788596]

Marcus, G. Neuroscience Fiction. 2012 Dec 2. Retrieved December 20 2012, from http://www.newyorker.com/online/blogs/newsdesk/2012/12/what-neuroscience-really-teaches-us-and-what-it-doesnt.html

McDonald RP. Structural Models and the Art of Approximation. Perspectives on Psychological Science. 2010; 5(6):675–686.10.1177/1745691610388766 [PubMed: 26161883]

McGue M, Osler M, Christensen K. Causal Inference and Observational Research: The Utility of Twins. Perspectives on Psychological Science. 2010; 5(5):546–556.10.1177/1745691610383511 [PubMed: 21593989]

McQueeny T, Schweinsburg BC, Schweinsburg AD, Jacobus J, Bava S, Frank LR, Tapert SF. Altered White Matter Integrity in Adolescent Binge Drinkers. Alcoholism-Clinical and Experimental Research. 2009; 33(7):1278–1285.10.1111/j.1530-0277.2009.00953.x

McTeague LM, Lang PJ, Laplante M-C, Cuthbert BN, Shumen JR, Bradley MM. Aversive Imagery in Posttraumatic Stress Disorder: Trauma Recurrence, Comorbidity, and Physiological Reactivity. Biological Psychiatry. 2010; 67(4):346–356. [PubMed: 19875104]

McTeague LM, Lang PJ, Laplante MC, Cuthbert BN, Strauss CC, Bradley MM. Fearful Imagery in Social Phobia: Generalization, Comorbidity, and Physiological Reactivity. Biological Psychiatry. 2009; 65(5):374–382. [PubMed: 18996510]

McTeague LM, Lang PJ, Wangelin BC, Laplante M-C, Bradley MM. Defensive mobilization in specific phobia: Fear specificity, negative sffectivity, and diagnostic prominence. Biological Psychiatry. 2012; 72(1):8–18. [PubMed: 22386377]

Medina KL, Schweinsburg AD, Cohen-Zion M, Nagel BJ, Tapert SF. Effects of alcohol and combined marijuana and alcohol use during adolescence on hippocampal volume and asymmetry. Neurotoxicol Teratol. 2007; 29(1):141–152. [PubMed: 17169528]

Meehl PE. Theoretical Risks and Tabular Asterisks - Karl, Ronald, and Slow Progress of Soft Psychology. Journal of Consulting and Clinical Psychology. 1978; 46(4):806–834.

Meehl PE. Apprasing and Amending Theories: The Strategy of Lakatosian Defense and Two Principles that Warrant It. Psychological Inquiry. 1990; 1(2):108–141.

Meehl PE. Factors and Taxa, Traits and Types, Differences of Degree and Differences in Kind. Journal of Personality. 1992; 60(1):117–174.

Meehl PE. Four Queries about Factor Reality. History and Philosophy of Psychology Bulletin. 1993; 5(2):4–5.

Meehl PE. Cliometric metatheory: II. Criteria scientists use in theory appraisal and why it is rational to do so. Psychological Reports. 2002; 91(2):339–404. [PubMed: 12416826]

Miller EK, Cohen JD. An integrative theory of prefrontal cortex function. Annual Review of Neuroscience. 2001; 24(1):167–202.10.1146/annurev.neuro.24.1.167

Miller GA, Rockstroh B. Endophenotypes in Psychopathology Research: Where Do We Stand? Annu Rev Clin Psychol. 201310.1146/annurev-clinpsy-050212-185540

Moffitt TE. The Neuropsychology of Conduct Disorder. Development and Psychopathology. 1993; 5(1-2):135–151.

Molina BS, Pelham WE Jr. Childhood predictors of adolescent substance use in a longitudinal study of children with ADHD. Journal of Abnormal Psychology. 2003; 112(3):497–507. [PubMed: 12943028]

Monroe SM, Harkness KL. Life stress, the" kindling" hypothesis, and the recurrence of depression: considerations from a life stress perspective. Psychological review. 2005; 112(2):417. [PubMed: 15783292]

Munafo MR, Durrant C, Lewis G, Flint J. Gene X environment interactions at the serotonin transporter locus. Biol Psychiatry. 2009; 65(3):211–219.10.1016/j.biopsych.2008.06.009 [PubMed: 18691701]

Muthén, LK.; Muthén, BO. Mplus user's guide. 6. Los Angeles, CA: Muthén & Muthén; 1998-2010.

Nagel BJ, Barlett VC, Schweinsburg AD, Tapert SF. Neuropsychological predictors of BOLD response during a spatial working memory task in adolescents: what can performance tell us about fMRI response patterns? J Clin Exp Neuropsychol. 2005; 27(7):823–839. [PubMed: 16183616]

Nagel BJ, Schweinsburg AD, Phan V, Tapert SF. Reduced hippocampal volume among adolescents with alcohol use disorders without psychiatric comorbidity. Psychiatry Res. 2005; 139(3):181–190. [PubMed: 16054344]

Najman JM, Behrens BC, Andersen M, Bor W, OCallaghan M, Williams GM. Impact of family type and family quality on child behavior problems: A longitudinal study. Journal of the American Academy of Child and Adolescent Psychiatry. 1997; 36(10):1357–1365. [PubMed: 9334548]

Nelson MR, Wegmann D, Ehm MG, Kessner D, St Jean P, Verzilli C, Mooser V, et al. An abundance of rare functional variants in 202 drug target genes sequenced in 14,002 people. Science. 2012; 337(6090):100–104.10.1126/science.1217876 [PubMed: 22604722]

NIMH. NIMH Research Domain Criteria (RDoC). 2011, 2012 Dec 21. Retrieved December 23, 2012, from http://www.nimh.nih.gov/research-funding/rdoc/nimh-research-domain-criteria-rdoc.shtml

O'Connor C, Rees G, Joffe H. Neuroscience in the Public Sphere. Neuron. 2012; 74(2):220–226. [PubMed: 22542177]

Olvet DM, Klein DN, Hajcak G. Depression symptom severity and error-related brain activity. Psychiatry Research. 2010; 179(1):30–37. [PubMed: 20630603]

Parker G, Fink M, Shorter E, Taylor MA, Akiskal H, Berrios G, Swartz C, et al. Issues for DSM-5: Whither melancholia? The case for its classification as a distinct mood disorder. The American Journal of Psychiatry. 2010; 167(7):745–747. [PubMed: 20595426]

Parvaz MA, Alia-Klein N, Woicik PA, Volkow ND, Goldstein RZ. Neuroimaging for drug addiction and related behaviors. Reviews in the Neurosciences. 2011; 22(6):609–624.10.1515/Rns.2011.055 [PubMed: 22117165]

Pashler H, Wagenmakers EJ. Editors' Introduction to the Special Section on Replicability in Psychological Science: A Crisis of Confidence? Perspectives on Psychological Science. 2012; 7(6):528–530.10.1177/1745691612465253 [PubMed: 26168108]

Patrick CJ, Bernat EM, Malone SM, Iacono WG, Krueger RF, McGue M. P300 amplitude as an indicator of externalizing in adolescent males. Psychophysiology. 2006; 43(1):84–92. [PubMed: 16629688]

Perlman G, Johnson W, Iacono WG. The heritability of P300 amplitude in 18-year-olds is robust to adolescent alcohol use. Psychophysiology. 2009; 46(5):962–969.10.1111/j.1469-8986.2009.00850.x [PubMed: 19558402]

Plomin R, DeFries JC, Loehlin JC. Genotype-environment interaction and correlation in the analysis of human behavior. Psychological Bulletin. 1977; 84:309–322. [PubMed: 557211]

Popper, K. The Logic of Scientific Discovery. New York: Basic Books; 1959.

Prisciandaro JJ, Roberts JE. A comparison of the predictive abilities of dimensional and categorical models of unipolar depression in the National Comorbidity Survey. Psychological Medicine. 2009; 39(07):1087–1096.10.1017/S0033291708004522 [PubMed: 18845012]

Pritchard JK. Are rare variants responsible for susceptibility to complex diseases? American Journal of Human Genetics. 2001; 69(1):124–137.10.1086/321272 [PubMed: 11404818]

Pritchard JK, Cox NJ. The allelic architecture of human disease genes: common disease-common variant…or not? Human Molecular Genetics. 2002; 11(20):2417–2423. [PubMed: 12351577]

Ravussin E, Valencia ME, Esparza J, Bennett PH, Schulz LO. Effects of a traditional lifestyle on obesity in Pima Indians. Diabetes Care. 1994; 17(9):1067–1074. [PubMed: 7988310]

Risch N. Interaction Between the Serotonin Transporter Gene (5-HTTLPR), Stressful Life Events, and Risk of Depression: A Meta-analysis (vol 301, pg 2462, 2009). Jama-Journal of the American Medical Association. 2009; 302(5):492–492.

Robinson TE, Berridge KC. The Neural Basis of Drug Craving - an Incentive-Sensitization Theory of Addiction. Brain Research Reviews. 1993; 18(3):247–291. [PubMed: 8401595]

Robinson TE, Gorny G, Mitton E, Kolb B. Cocaine self-administration alters the morphology of dendrites and dendritic spines in the nucleus accumbens and neocortex. Synapse. 2001; 39(3): 257–266. [PubMed: 11169774]

Rubin DB. Estimating the causal effects of smoking. Statistics in Medicine. 2001; 20(9-10):1395–1414. [PubMed: 11343361]

Rubin DB. The design versus the analysis of observational studies for causal effects: Parallels with the design of randomized trials. Statistics in Medicine. 2007; 26(1):20–36.10.1002/Sim.2739 [PubMed: 17072897]

Rubin DB. For Objective Causal Inference, Design Trumps Analysis. Annals of Applied Statistics. 2008; 2(3):808–840.10.1214/08-Aoas187

Ruchsow M, Grön G, Reuter K, Spitzer M, Hermle L, Kiefer M. Error-related brain activity in patients with obsessive-compulsive disorder and in healthy controls. Journal of Psychophysiology. 2005; 19(4):298–304.10.1027/0269-8803.19.4.298

Ruchsow M, Herrnberger B, Wiesend C, Grön G, Spitzer M, Kiefer M. The effect of erroneous responses on response monitoring in patients with major depressive disorder: A study with event-related potentials. Psychophysiology. 2004; 41(6):833–840.10.1111/j.1469-8986.2004.00237.x [PubMed: 15563336]

Ruchsow M, Herrnberger Br, Beschoner P, Grön G, Spitzer M, Kiefer M. Error processing in major depressive disorder: Evidence from event-related potentials. Journal of Psychiatric Research. 2006; 40(1):37–46. [PubMed: 15882872]

Rutter M. Developmental catch-up, and deficit, following adoption after severe global early privation. English and Romanian Adoptees (ERA) Study Team. J Child Psychol Psychiatry. 1998; 39(4): 465–476. [PubMed: 9599775]

Rutter M. Proceeding From Observed Correlation to Causal Inference The Use of Natural Experiments. Perspectives on Psychological Science. 2007; 2(4):377–395.10.1111/j.1745-6916.2007.00050.x [PubMed: 26151974]

Sartor CE, Lynskey MT, Heath AC, Jacob T, True W. The role of childhood risk factors in initiation of alcohol use and progression to alcohol dependence. Addiction. 2007; 102(2):216–225. [PubMed: 17222275]

Scarr S, McCartney K. How people make their own environments: A theory of genotype => environment effects. Child Development. 1983; 54:424–435. [PubMed: 6683622]

Schizophrenia Working Group of the Psychiatric Genomics Consortium. Biological insights from 108 schizophrenia-associated genetic loci. Nature. 2014; 511(7510):421–427. [PubMed: 25056061]

Schrijvers D, de Bruijn ERA, Maas Y, De Grave C, Sabbe BGC, Hulstijn W. Action monitoring in major depressive disorder with psychomotor retardation. Cortex. 2008; 44(5):569–579. [PubMed: 18387589]

Schrijvers D, De Bruijn ERA, Maas YJ, Vancoillie P, Hulstijn W, Sabbe BGC. Action monitoring and depressive symptom reduction in major depressive disorder. International Journal of Psychophysiology. 2009; 71(3):218–224. [PubMed: 18926863]

Shadish, WR.; Cook, TD.; Campbell, DT. Experimental and quasi-experimental designs for generalized causal inference. Boston: Houghton Mifflin; 2001.

Simpson HB, Nee JC, Endicott J. First-episode major depression. Few sex differences in course. Archives of General Psychiatry. 1997; 54(7):633–639. [PubMed: 9236547]

Slade T, Watson D. The structure of common DSM-IV and ICD-10 mental disorders in the Australian general population. Psychological Medicine. 2006; 36(11):1593–1600.10.1017/s0033291706008452 [PubMed: 16882356]

Smith GD. Mendelian randomization for strengthening causal inference in observational studies: Application to gene X environment interactions (vol 5, pg 527, 2010). Perspectives on Psychological Science. 2011; 6(3):314–314.10.1177/1745691611411273 [PubMed: 26168522]

Solomon ARI, Haaga DAF, Arnow BA. Is Clinical Depression Distinct from Subthreshold Depressive Symptoms? A Review of the Continuity Issue in Depression Research. The Journal of Nervous and Mental Disease. 2001; 189(8):498–506. [PubMed: 11531201]

Spear LP. The adolescent brain and age-related behavioral manifestations. Neuroscience and Biobehavioral Reviews. 2000; 24(4):417–463. [PubMed: 10817843]

Spearman C. “General intelligence” objectively determined and measured. American Journal of Psychology. 1904; 15:201–292.

Speliotes EK, Willer CJ, Berndt SI, Monda KL, Thorleifsson G, Jackson AU, Magic, et al. Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. Nature Genetics. 2010; 42(11):937–U953.10.1038/ng.686 [PubMed: 20935630]

Sterba SK, Pek J. Individual Influence on Model Selection. Psychological Methods. 201210.1037/a0029253

Stern ER, Liu Y, Gehring WJ, Lister JJ, Yin G, Zhang J, Taylor SF, et al. Chronic medication does not affect hyperactive error responses in obsessive-compulsive disorder. Psychophysiology. 2010; 47(5):913–920.10.1111/j.1469-8986.2010.00988.x [PubMed: 20230510]

Stroud CB, Davila J, Moyer A. The relationship between stress and depression in first onsets versus recurrences: A meta-analytic review. Journal of Abnormal Psychology. 2008; 117(1):206. [PubMed: 18266498]

Sullivan PF, Daly MJ, O’Donovan M. Genetic architectures of psychiatric disorders: the emerging picture and its implications. Nature Reviews Genetics. 2012; 13(8):537–551.10.1038/nrg3240

Sullivan PF, Neale MC, Kendler KS. Genetic epidemiology of major depression: Review and meta-analysis. The American Journal of Psychiatry. 2000; 157(10):1552–1562.10.1176/appi.ajp.157.10.1552 [PubMed: 11007705]

Taleb, NN. The black swan: The impact of the highly improbable. Random House; 2010.

Tavris, C. Psychobabble and Biobunk: Using Psychological Science to Think Critically About Popular Psychology. Pearson; 2010.

Thyreau B, Schwartz Y, Thirion B, Frouin V, Loth E, Vollstadt-Klein S, Poline J-B, et al. Very large fMRI study using the IMAGEN database: Sensitivity-specificity and population effect modeling in relation to the underlying anatomy. NeuroImage. 2012; 61(1):295–303. [PubMed: 22425669]

Tukey, JW. Exploratory Data Analysis. Reading, MA: 1977.

Tukey JW. We Need Both Exploratory and Confirmatory. American Statistician. 1980; 34(1):23–25.10.2307/2682991

Vaidyanathan U, Isen JD, Malone SM, Miller MB, McGue M, Iacono WG. Heritability and molecular genetic basis of electrodermal activity: A genome-wide association study. Psychophysiology. 2014; 51(12):1259–1271. [PubMed: 25387706]

Vaidyanathan U, Malone SM, Donnelly JM, Hammer MA, Miller MB, McGue M, Iacono WG. Heritability and molecular genetic basis of antisaccade eye tracking error rate: A genome-wide association study. Psychophysiology. 2014; 51(12):1272–1284. [PubMed: 25387707]

Vaidyanathan U, Malone SM, Miller MB, McGue M, Iacono WG. Heritability and molecular genetic basis of acoustic startle eye blink and affectively modulated startle response: A genome-wide association study. Psychophysiology. 2014; 51(12):1285–1299. [PubMed: 25387708]

Vaidyanathan U, Welo EJ, Malone SM, Burwell SJ, Iacono WG. The effects of recurrent episodes of depression on startle responses. Psychophysiology. 2014; 51(1):103–109.10.1111/psyp.12152 [PubMed: 24117611]

Vanyukov MM, Tarter RE, Kirillova GP, Kirisci L, Reynolds MD, Kreek MJ, Ridenour TA, et al. Common liability to addiction and “gateway hypothesis”: Theoretical, empirical and evolutionary perspective. Drug and Alcohol Dependence. 2012; 123(Suppl 1):S3–S17.10.1016/j.drugalcdep.2011.12.018 [PubMed: 22261179]

Visscher PM, Brown MA, McCarthy MI, Yang J. Five years of GWAS discovery. American Journal of Human Genetics. 2012; 90(1):7–24.10.1016/j.ajhg.2011.11.029 [PubMed: 22243964]

Volkow ND, Mullani N, Gould L, Adler SS, Guynn RW, Overall JE, Dewey S. Effects of Acute Alcohol-Intoxication on Cerebral Blood-Flow Measured with Pet. Psychiatry Research. 1988; 24(2):201–209. [PubMed: 3261427]

Vollebergh WA, Iedema J, Bijl RV, de Graaf R, Smit F, Ormel J. The structure and stability of common mental disorders: the NEMESIS study. Archives of General Psychiatry. 2001; 58(6): 597–603. [PubMed: 11386990]

Voss M. How to spot pseudoneuroscience and biobunk. Observer. 2012; 25:36–37.

Vrana SR, Spence EL, Lang PJ. The startle probe response: A new measure of emotion? Journal of Abnormal Psychology. 1988; 97(4):487–491.10.1037/0021-843x.97.4.487 [PubMed: 3204235]

Vrieze SI, Hicks BM, Iacono WG, McGue M. Decline in Genetic Influence on the Co-Occurrence of Alcohol, Marijuana, and Nicotine Dependence Symptoms From Age 14 to 29. American Journal of Psychiatry. 201210.1176/appi.ajp.2012.11081268

Vrieze SI, Iacono WG, McGue M. Confluence of genes, environment, development, and behavior in a post Genome-Wide Association Study world. Development and Psychopathology. 2012; 24(4): 1195–1214.10.1017/S0954579412000648 [PubMed: 23062291]

Vrieze SI, Malone SM, Pankratz N, Vaidyanathan U, Miller MB, Kang HM, Iacono WG, et al. Genetic associations of nonsynonymous exonic variants with psychophysiological endophenotypes. Psychophysiology. 2014; 51(12):1300–1308. [PubMed: 25387709]

Vrieze SI, Malone SM, Vaidyanathan U, Kwong A, Kang HM, Zhan X, Locke AE, et al. In search of rare variants: Preliminary results from whole genome sequencing of 1,325 individuals with psychophysiological endophenotypes. Psychophysiology. 2014; 51(12):1309–1320. [PubMed: 25387710]

Vrieze SI, McGue M, Iacono WG. The interplay of genes and adolescent development in substance use disorders: leveraging findings from GWAS meta-analyses to test developmental hypotheses about nicotine consumption. Human Genetics. 2012; 131(6):791–801. [PubMed: 22492059]

Vrieze SI, McGue M, Miller MB, Iacono WG. Three mutually informative ways to understand the genetic relationships among behavioral disinhibition, alcohol use, drug use, nicotine use, and their co-occurrence: Twin biometry, GCTA, and genome-wide scoring. 2012 Manuscript submitted for publication.

Vul E, Harris C, Winkielman P, Pashler H. Puzzlingly high correlations in fMRI studies of emotion, personality, and social cognition. Perspectives on Psychological Science. 2009; 4(3):274–290.10.1111/j.1745-6924.2009.01125.x [PubMed: 26158964]

Warner V, Weissman MM, Fendrich M, Wickramaratne P, Moreau D. The course of major depression in the offspring of depressed parents. Incidence, recurrence, and recovery. Archives of General Psychiatry. 1992; 49(10):795–801. [PubMed: 1417432]

Weinberg A, Klein DN, Hajcak G. Increased error-related brain activity distinguishes generalized anxiety disorder with and without comorbid major depressive disorder. Journal of Abnormal Psychology. 2012; 121(4):885–896.10.1037/a0028270 [PubMed: 22564180]

Weinberg A, Olvet DM, Hajcak G. Increased error-related brain activity in generalized anxiety disorder. Biological Psychology. 2010; 85(3):472–480. [PubMed: 20883743]

Wichers M, Geschwind N, van Os J, Peeters F. Scars in depression: is a conceptual shift necessary to solve the puzzle? Psychological Medicine. 2010; 40(03):359–365.10.1017/S0033291709990420 [PubMed: 20120516]

Wilson S, DiRago AC, Iacono WG. Prospective inter-relationships between late adolescent personality and major depressive disorder in early adulthood. Psychological Medicine. 2014; 44(03):567–577. [PubMed: 23689064]

Wilson S, Vaidyanathan U, Miller MB, McGue M, Iacono WG. Premorbid risk factors for major depressive disorder: Are they associated with early onset and recurrent course? Development and Psychopathology. 2014; 26(4pt2):1477–1493. [PubMed: 25422974]

Yang YH. Can the strengths of AIC and BIC be shared? A conflict between model indentification and regression estimation. Biometrika. 2005; 92(4):937–950.

Yoon HH, Iacono WG, Malone SM, McGue M. Using the brain P300 response to identify novel phenotypes reflecting genetic vulnerability for adolescent substance misuse. Addictive Behaviors. 2006; 31(6):1067–1087.10.1016/j.addbeh.2006.03.036 [PubMed: 16644137]

Young SE, Stallings MC, Corley RP, Krauter KS, Hewitt JK. Genetic and environmental influences on behavioral disinhibition. Amercan Journal of Medical Genetics. 2000; 96(5):684–695.10.1002/1096-8628(20001009)96:5<684: :AID-AJMG16>3.0.CO;2-G

Yuan Z, Yang YH. Combining linear regression models: When and how? Journal of the American Statistical Association. 2005; 100(472):1202–1214.10.1198/016214505000000088

Zucker RA, Heitzeg MM, Nigg JT. Parsing the Undercontrol/Disinhibition Pathway to Substance Use Disorders: A Multilevel Developmental Problem. Child Development Perspectives. 2011; 5(4): 248–255.10.1111/j.1750-8606.2011.00172.x [PubMed: 22116786]