The Research Domain Criteria (RDoC): An analysis of methodological and conceptual challenges

Scott O. Lilienfeld*

Department of Psychology, Emory University, United States

A B S T R A C T

In a bold effort to address the longstanding shortcomings of the Diagnostic and Statistical Manual (DSM) framework for the classification and diagnosis of psychopathology, the National Institute of Mental Health recently launched a research program—the Research Domain Criteria (RDoC)—in the hopes of developing an alternative taxonomic system rooted in dysfunctional brain circuitry. Although the RDoC endeavor has considerable promise, it faces several methodological and conceptual challenges, four of which I address here: (a) an overemphasis on biological units and measures, (b) neglect of measurement error, (c) biological and psychometric limitations of endophenotypes, and (d) the distinction between biological predispositions and their behavioral manifestations. Because none of these challenges is in principle insurmountable, I encourage investigators to consider potential remedies for them. RDoC is a calculated gamble that appears to be worth the risk, but its chances of success will be maximized by a thoughtful consideration of hard-won lessons learned—but frequently forgotten—over the past several decades of psychological and psychiatric research.

© 2014 Elsevier Ltd. All rights reserved.

The Research Domain Criteria (RDoC): an analysis of conceptual and methodological challenges

It perhaps goes without saying that our current model of psychiatric classification has left many researchers and clinicians dissatisfied (Frances, 2013; Greenberg, 2013). In contrast to contemporary systems of classification in internal medicine, the Diagnostic and Statistical Manual of Mental Disorders (DSM), which first appeared in 1952, diagnoses individuals almost exclusively on the basis of their signs (overt manifestations) and symptoms (subjective reports) rather than on the basis of their pathology and etiology. As a consequence, the DSM may erroneously place individuals who share superficial similarities but whose pathology springs from different sources into the same diagnostic category. In many ways, psychiatric diagnosis and treatment circa 2014 is akin to much of medical diagnosis and treatment 150 years ago. For example, whereas physicians once diagnosed and treated all forms of fever similarly, they now recognize that fever is merely a nonspecific manifestation of a plethora of underlying maladies, each of which necessitates a different intervention (Kihlstrom, 2002).

The recent release of the fifth edition of the DSM (DSM-5; American Psychiatric Association, 2013) brought debates regarding psychiatric classification into bold relief, with some (e.g., Kupfer & Regier, 2010)largely defending the status quo and others (Whooley & Horwitz, 2013)forcefully criticizing it. Hence, it is not surprising that these disputes renewed calls for a competing framework to the DSM, ideally one tied more closely to pathology and etiology.

The DSM and its discontents

Of course, these disagreements are hardly new. Deep-seated discontent regarding the prevailing DSM model has been a recurring theme in clinical psychology and psychiatry over the past several decades (e.g., Faust & Miner, 1986; Kirk & Kutchins, 1992; Widiger & Clark, 2000). Indeed, both apologists and critics of the DSM have apparently managed to achieve consensus on one point: The current system of classification and diagnosis is far from optimal. Even the most fervent defenders of the DSM readily concede that it has fallen short of its goal of carving nature at its joints. There is no need to reiterate all of the familiar shortcomings with the DSM here, but several persistent problems are worth highlighting (see Lilienfeld, Smith, & Watts, 2013, for a review).
Extensive “comorbidity” (but see Lilienfeld, Waldman, & Israel, 1994, for a critique of this concept and term as commonly applied in psychopathology research). There is extensive co-occurrence and, even more troubling, covariation, among many putatively separable psychological conditions, suggesting that these conditions are often slightly different variants of shared etiological processes (Cramer, Waldrop, van der Maas, & Borsboom, 2010; Vaidyanathan, Patrick, & Iacono, 2011). For many DSM disorders, such as posttraumatic stress disorder (PTSD; Brady, Kline, Brewerton, & Kaplun, 2000), childhood externalizing and internalizing disorders, and all personality disorders (Grove & Tellegen, 1991), comorbidity — in the sense of co-occurrence — is the rule rather than the exception, with the substantial majority of individuals with a given condition meeting criteria for one or more additional conditions (Lilienfeld, 2007). In an especially extreme case, one patient in a published study met diagnostic criteria for all ten DSM personality disorders (Widiger et al., 1998).

Unsupported imposition of a categorical measurement model. Data increasingly suggest that most DSM disorders, including the majority of mood, anxiety, eating, and personality disorders, are dimensional rather than taxonic (categorical) in nature (Haslam, Holland, & Kuppens, 2012). The DSM’s imposition of a categorical measurement model on these conditions therefore runs afoul of the best available scientific evidence. Even setting aside the ontological status of DSM disorders, a categorical model decreases reliability and validity relative to a dimensional model by omitting valuable psychometric information (Markon, Chmielewski, & Miller, 2011).

Inadequate construct validity. Although many DSM-5 diagnoses clearly possess at least some construct validity (cf., Greenberg, 2013; Insel, 2013), the capacity of these diagnoses to statistically predict important external criteria, such as natural history, family history, performance on laboratory variables (Robins & Guze, 1970) and treatment response (Kendler, 1990), has often been disappointing. For example, the DSM diagnosis of antisocial personality disorder is largely unrelated to laboratory markers of emotional processing or decision-making, and is only a modest marker of criminal recidivism (Risher & Kosson, 2013).

Heterogeneity. On a related front, many DSM diagnoses are almost certainly heterogeneous constellations of features in multidimensional space. The symptomatic heterogeneity of these conditions is in part a consequence of the polythetic (“Chinese menu”) algorithms used to derive DSM diagnoses, which permit large numbers of different sign and symptom combinations to qualify for the same disorder. For example, in DSM-5, there are 636,120 ways to meet diagnostic criteria for PTSD (Galatzer-Levy & Bryant, 2013). Nevertheless, the problem extends to etiological heterogeneity as well. The DSM-5 diagnosis of alcohol use disorder (which comprised alcohol dependence and alcohol abuse in DSM-IV), for instance, appears to subsume several largely distinct subtypes that differ substantially in age of onset, family history, and covariation with other psychiatric conditions (Moss, Chen, & Yi, 2007).

Not otherwise specified (NOS) diagnoses. For most major classes of DSM psychopathology, including personality and eating disorders, the modal diagnosis is NOS, meaning that most patients with mental disorders do not fit into any extant category (Westen, 2012). This finding suggests that the DSM has failed to achieve an ideal classification system, which contains few or no intermediate cases (Frances, 1980; Lilienfeld, VanValenkean, Larrntz, & Akiskal, 1986).

Despite the heroic efforts of multiple DSM task forces, these and other shortcomings have endured across multiple editions of the manual. Indeed, the often vitriolic controversies surrounding the 2013 release of DSM-5 (e.g., Spitzer & Frances, 2010) may have obscured a crucial truth: most of the revisions from DSM-III (American Psychiatric Association, 1980) to DSM-5 were in fact relatively minor, and all of the principal structural drawbacks plaguing earlier editions of the DSM remain. The same problems afflict the classification system in the International Classification of Diseases. (ICD), which overlaps with the DSM (Frances, 2014). We remain saddled with a diagnostic paradigm that (a) relies almost exclusively on superficial signs and symptoms and (b) is largely divorced from pathology and etiology. The current classification system, although reasonably reliable, is almost exclusively descriptive rather than explanatory (McHugh & Slavney, 2011). As a consequence, it seems unlikely to map well onto the genuine causal structure of psychopathology.

The failure of the DSM and ICD to generate a model of classification based on pathology and etiology cannot be laid entirely at the feet of the developers and shapers of these manuals. After all, our field’s understanding of the causes of most mental disorders remains in its infancy. Hence, a theory-neutral taxonomy may be the best we can manage given our present state of knowledge (Wakefield, 1999). Moreover, efforts to develop etiologically-informed alternatives to the DSM, such as psychodynamic (Alliance of Psychoanalytic Associations, 2006) or behavior analytic (Follette & Houts, 1996) classification systems, have gained minimal scientific traction, probably because their research base is insufficient to support the edifice upon which they are constructed. Unless and until a better alternative comes along, we appear to be stuck with the DSM and ICD, whether we like it or not.

The Research Domain Criteria (RDoC) proposal: mental disorders as dysfunctions of brain circuits

Against the backdrop of lingering discontent with the DSM and ICD, in 2009 the National Institute of Mental Health (NIMH) initiated a bold initiative to transform the current framework of psychiatric classification and diagnosis into an explicitly biological system (Cuthbert, 2014; Insel et al., 2010; Sanislow et al., 2010). Dubbed the Research Domain Criteria (RDoC), largely as an homage to the Research Diagnostic Criteria (RDC) that were precursors to the DSM-III criteria (Spitzer, Endicott, & Robins, 1978), this new endeavor is intended to inaugurate a paradigm shift. Rather than base psychiatric diagnosis on presenting signs and symptoms, as do the DSM and ICD, RDoC strives to anchor psychiatric classification and diagnosis in a scientifically supported model of neural circuitry. Specifically, RDoC conceptualizes mental disorders as dysfunctions in brain systems that bear important adaptive implications, such as systems linked to reward responsiveness and threat sensitivity (see also Harkness, Reynolds, & Lilienfeld, 2013).

At least for the foreseeable future, RDoC is not envisioned as a system of psychiatric classification and diagnosis in its own right. Nor does it adopt an a priori stance on what form such a system should eventually assume (Insel, 2014). Instead, RDoC is conceptualized as a long-term program of research that may ultimately yield the broad outlines of such a system (MacDonald & Krueger, 2013). In the words of the current NIMH director and his colleagues, RDoC is “a vision for the future” (Insel et al., 2010, p. 749) rather than a full-fledged proposal. As of this writing, RDoC is largely fluid in its mission. Such flexibility may well be justified in

---

1. For each major disorder class, DSM-5 now subdivides the NOS category, often pejoratively called the “wastebasket” category, into “other specified” and “unspecified” diagnoses.
the early phases of scientific investigation, in which hypothesis generation (“the context of discovery”) should often take precedence over hypothesis testing (“the context of justification”) (Kell & Oliver, 2004).

Still, the RDoC research program is by no means entirely open-ended. RDoC provides researchers with an explicit rubric for guiding their investigations. As can be seen in Fig. 1, RDoC proposes that research efforts be conducted within a two dimensional matrix, which is intended to serve as provisional guide for research purposes (Morris & Cuthbert, 2012; Weinberger & Goldberg, 2014). On the horizontal axis lie seven units of analysis organized roughly from more to less “basic”: genes, molecules, cells, circuits, physiology, behavior, and self-reports (the matrix also includes a column for “paradigms,” allowing investigators to indicate which tasks are especially useful for the research question at hand). On the vertical axis lie five broad psychobiological domains/constructs that correspond to brain-based circuits that are relevant to psychopathology: negative valence systems (e.g., threat, loss), positive valence systems (e.g., approach motivation, responsiveness to reward); cognitive systems (e.g., attention, working memory), social processes (e.g., theory of mind, dominance), and arousal/regulatory systems. This matrix is not exhaustive, but it affords a reasonably comprehensive heuristic for grounding research projects in promising indicators of psychobiological systems. It is perhaps unfortunate, however, that the proposed RDoC brain systems are not aligned more explicitly with well-established dimensions of normal and abnormal personality that bear clear-cut adaptive implications, such as the Personality Psychopathology Five (PSY-5; Harkness, Finn, McNulty, & Shields, 2012; Harkness et al., 2013).

Philosophically and methodologically, RDoC rests on several bedrock assumptions (Cuthbert & Insel, 2013). RDoC is strongly translational in emphasis, encouraging researchers to apply the basic science of brain systems and behavior to an understanding of the correlates and causes of mental disorders. RDoC also encourages a dimensional framework for psychopathology in light of evidence that the activity of most brain circuits, such as reward and threat systems, is continuously distributed, with few or no clear-cut boundaries demarcating normality from abnormality. In addition, RDoC strives to accord roughly equal weight to different levels of analysis, including the biological and behavioral (Cuthbert & Insel, 2013). Consistent with the view of science as a self-correcting enterprise (Herbert et al., 2000), RDoC is provisional and open to revision. As a consequence, novel brain-based constructs and behavioral units of analysis may be added to the matrix with the emergence of new neuroscience and social evidence.

There is much to admire about the RDoC proposal. Among other things, it has the potential to loosen the longstanding hegemony of the DSM system over research, which has stifled investigators’ capacity to explore scientifically promising alternatives to the status quo model of psychiatric classification (see also Berenbaum, 2013; Harkness & Lilienfeld, 2013; Markon, 2013, for discussions). Moreover, RDoC is broadly consistent with a dimensional approach to psychopathology, which accords with the bulk of evidence derived from taxometric studies (Harkness & Lilienfeld, 1997; Haslam et al., 2012). At the same time, RDoC allows for the possibility of threshold effects (“tipping points”) or other categorical effects (Cuthbert & Insel, 2013), whereby certain psychopathological phenomena differ qualitatively rather than quantitatively from normality. Finally, RDoC promises to capitalize on the burgeoning corpus of knowledge concerning clinical neuroscience by applying it to the classification and etiology of psychopathology. In all of these respects, RDoC appears to be a welcome development, especially given that the DSM’s scientific yield, which has certainly been substantial despite its noteworthy defects (Regier, Narrow, Kuhl, & Kupfer, 2009), seems to be approaching a plateau.

At the same time, RDoC brings with it a number of conceptual and methodological challenges, most or all of which have received insufficient attention (e.g., Sanislow et al., 2010). None of these

<table>
<thead>
<tr>
<th>DOMAINS/CONSTRUCTS</th>
<th>Genes</th>
<th>Molecules</th>
<th>Cells</th>
<th>Circuits</th>
<th>Physiology</th>
<th>Behavior</th>
<th>Self-Reports</th>
<th>Paradigms</th>
</tr>
</thead>
<tbody>
<tr>
<td>Negative Valence Systems</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Acute threat (“fear”)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Potential threat (“anxiety”)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sustained threat</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Loss</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Frustrative nonreward</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Positive Valence Systems</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Approach motivation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Initial responsiveness to reward</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sustained responsiveness to reward</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reward learning</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Habit</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cognitive Systems</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Attention</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Perception</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Working memory</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Declarative memory</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Language behavior</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cognitive (effortful) control</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Systems for Social Processes</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Affiliation/attachment</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Social communication</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Perception/understanding of self</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Perception/understanding of others</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Arousal/Modulatory Systems</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Arousal</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Biological rhythms</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sleep-wake</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Fig. 1. The current matrix for the Research Domain Criteria (RDoC). From Cuthbert (2014). Permission pending.
challenges is new. To the contrary, they have been recognized in various guises over many decades of psychological and psychiatric research, yet they appear not to have accorded extensive discussion in the articles or public fora concerning RDoC. In the remainder of this article, I address four overarching challenges that, unless considered thoughtfully and addressed, may hinder RDoC's scientific progress. Several of these challenges overlap, but I have separated them for the purposes of exposition.

Because I believe that the RDoC mission has potential merit, my intent is constructive; hence, for each challenge, I offer tangible recommendations. None of these challenges, I contend, is insurmountable. To the contrary, a more thoughtful consideration of each challenge should help to strengthen rather than weaken RDoC's scientific potential. By bringing these challenges to the fore and proposing potential remedies for them, I hope to help maximize the chances that RDoC will bear long-term scientific fruit.

**Challenge #1: overemphasis on biological units and measures**

The first major conceptual challenge to be confronted is RDoC's assumption that mental disorders are fundamentally "disorders of brain circuits" (Insel et al., 2010, p. 749). As a corollary, RDoC presumes that "the tools of clinical neuroscience ... can be used to identify dysfunction in neural circuits" (Morris & Cuthbert, 2012, p. 33). Of course, at some level, this assumption is necessarily true given that all psychological disorders — indeed, all psychological phenomena — are mediated by the brain and the remainder of the central nervous system (Kendler, 2005). Yet investigators operating within the RDoC framework must be careful not to confuse biological mediation with biological etiology.

This tempting semantic slippage is subtle but substantial in its implications. For example, in principle, a psychological condition could be triggered largely by psychosocial factors, such as childhood sexual or physical abuse. Although this condition would of course be mediated by brain circuitry, its etiology would be primarily environmental. The danger here is that unwary investigators seeking to elucidate the causes of this condition might direct their etiological efforts away from their principal source, in this case early abuse, and incorrectly toward brain-based causes, such as volumetric abnormalities in the size of subcortical regions or connectivity differences across regions.

To be fair, RDoC does not explicitly commit the logical error of confusing biological mediation with biological etiology. Nevertheless, it may inadvertently foster this error by de-emphasizing psychosocial variables in its conceptualization. The current RDoC matrix focuses almost exclusively on intra-individual variables, with little or no explicit coverage of extra-individual variables, such as the social or cultural context (Berenbaum, 2013; Whooley, 2014; Haffel et al., 2008; Kwapil, 1998) than any available biological measure." (p. 933). If so, RDoC should not peremptorily exclude the possibility that self-report measures alone may in some cases be the best indicators of relevant biological systems. Moreover, although RDoC should certainly encourage research that draws on measures drawn from multiple units of analysis, it should not adopt an a priori stance on which measures will prove most valid for detecting individual differences in neural circuitry.

A related source of logical confusion is the erroneous equation of biological units of analysis with biological measures of these units. For example, in a comment supporting NIMH Director Thomas Insel's RDoC mission, John Scully, the American Psychiatric Association's chief officer, stated that "We want him [Insel] to get biomarkers for us" (Gever, 2013). Similarly, David Kupfer, co-chair of DSM-5, said that "While we don't yet have the biomarkers that we are hoping on the edge of discovery, patients can't keep waiting, and we can't keep waiting" (Gever, 2013). Both comments seem to imply that indicators that are intrinsically biological in nature ("biomarkers"), such as neurotransmitter metabolites or neuroimaging data, are necessary to provide valid measures of biological systems. If so, this conclusion is unwarranted, because self-report and behavioral data can, and often do, provide valid indicators of such systems.

**Recommendations**

Investigators working within the RDoC framework should be certain not to accord potential psychosocial variables, including situational, cultural, and developmental variables, short shrift. They should also not assume that measures at one level of analysis are necessarily best suited for detecting individual differences in that unit of analysis. For example, it is entirely possible that certain self-report indicators (e.g., questionnaire measures of harmavoidance) will prove to be most valid for detecting individual differences in certain biological systems (e.g., threat sensitivity), or conversely,
that certain psychophysiological indicators (e.g., left hemisphere activation) will prove to be most valid for detecting individual differences in personality trait dimensions (e.g., dispositional positive emotionality; e.g., Coan & Allen, 2004; Davidson, 1992). RDoC’s stance in this regard should be purely empirical, and should not be dictated by a priori assumptions that are not necessarily rooted in data, such as assumptions that self-report indices will necessarily be inferior to biological indices in the detection of individual differences in biological systems.

Challenge #2: neglect of measurement error

With the advent of RDoC, more research emphasis will almost certainly be accorded to putative laboratory (including both behavioral and psychophysiological) indicators of biological processes, such as fear-potentiated startle to detect threat sensitivity (Davis, 2006; Lang, 1995), go-no go responding tasks to detect passive avoidance learning deficits (Newman & Kosson, 1986), and delay discounting tasks to detect sensitivity to short-term rewards (Bickel, Yi, Landes, Hill, & Baxter, 2011). To the extent that such measures serve as counterweights to research psychology’s prevailing neglect of observed behavior in favor of more easily collected self-reports (Baumeister, Vohs, & Funder, 2007), this change in emphasis should enhance the field’s breath of coverage and ideally, its external validity.

Consistent with the RDoC’s call, several prominent authors have long argued that psychiatric diagnosis should begin to transition to a laboratory-based approach (see Widiger & Clark, 2000, for a review). For example, Kihlstrom (2002) suggested that descriptive psychology should move beyond the neo-Kraepelianian framework (Blashfield, 1984) that inspired DSM-III (American Psychiatric Association, 1980) and subsequent versions of the manual, replacing signs and symptoms with well-validated laboratory measures. For Kihlstrom, laboratory-based psychiatric diagnoses will be needed to bring psychiatry in line with the rest of medicine (see also Nemeroft, Kilts, & Berns, 1999).

Kihlstrom’s (2002) ambitious vision, echoed by RDoC, is worth pursuing. At the same time, laboratory measures are often associated with unappreciated psychometric weaknesses, a crucial caveat neglected in virtually all RDoC documents (e.g., Sanislow et al., 2010). As Epstein (1979, 1980) observed over three decades ago, psychologists have long granted laboratory measures an undeserved scientific cachet, often “giving them a pass” with respect to fundamental psychometric requirements (see also Tryon, 1973). Indeed, psychologists have long recognized that laboratory measures often display low levels of temporal and cross-situational consistency, largely because they contain substantial components of “situational uniqueness.” Specifically, performance on such measures can be affected by a plethora of contextual and situational factors, including the mood and alertness of the participant, the time of day, the experimental instructions, the nature of the laboratory setting, the perceived attitude of the experimenter, and perhaps most important, the particulars of the laboratory paradigm itself (Epstein, 1979; Kendler & Neale, 2010). With respect to the lattermost source of variance, Mischel (1968) famously observed that even seemingly trivial changes in experimental paradigms can lead to dramatic changes in these measures’ intercorrelations and associations with other measures, an overarching conclusion that has stood the test of time (Kenrick & Funder, 1988). In this context, Epstein (1979) asked rhetorically:

Can it be that overevaluation of the experimental method, as normally practiced, blinded researchers to the inherent limitations of studying behavior on single occasions? Given the awe in which laboratory experimental procedures have been held, who would have dared to think that they often fail to meet one of the most fundamental scientific tests of all, temporal reliability (replicability)? (p. 1121; emphasis added).

Similarly, Block (1977; see also Mischel & Peake, 1982; Tellegen, 1991) noted that T data (test data), that is, data drawn from isolated laboratory indicators, are almost always more unreliable and erratic in their relations with (a) both each other and (b) other measures compared with S data (self-report data) and R data (rating data). S and R data, although possessing psychometric limitations of their own (e.g., susceptibility to response biases), are typically aggregated across multiple diverse situations. In this way, they can minimize situational error and thereby yield more reliable and construct valid composites of behavior across multiple situations (Epstein, 1980; Rushton, Brainard, & Pressley, 1983). In contrast, T data are rarely aggregated across diverse situations, only across multiple trials of the same measure (Epstein, 1983).

One may wonder whether the often unwarranted scientific legitimacy accorded to laboratory measures may stem in part from a representativeness heuristic (Tversky & Kahneman, 1974) whereby psychological measures that resemble those in the “harder,” laboratory sciences are automatically presumed to be especially rigorous. As Schweitzer et al. (2011) described this heuristic in the case of psychological research, “The better a science fits your stereotypes, the better the science” (p. 363). This undeserved prestige may be especially true for neuroimaging measures given that neuroscience explanations often strike people as inherently more convincing than non-neuroscience explanations (Weisberg, Keil, Goodstein, Rawson, & Gray, 2008; but see Farah & Hook, 2013, for a critique of the “neuroseduction” literature).

For example, few investigators have examined the test–retest reliability of measures of functional magnetic resonance imaging (fMRI; Bennett & Miller, 2010), even though test–retest reliability is a fundamental expectation of measures in other psychological domains. In an analysis of 63 studies, Bennett and Miller (2010) found that the test–retest reliability of fMRI measures was typically modest, with intraclass correlations (ICCs) averaging 0.50 (see also Kendler & Neale, 2010; Vul, Harris, Wexkelman, & Pashler, 2009, for similar conclusions). The reliabilities varied markedly across studies, with some values being considerably higher and others considerably lower. Furthermore, the average cluster overlap value for voxels in Bennett and Millers’s analysis was 29%, meaning that only 29% of voxels that were statistically significant in one study were significant in a second study. Although test–retest reliabilities for fMRI data tend to be higher with briefer intervals, even back-to-back scans (taken within one hour) exhibited an average cluster overlap of only 33% (Bennett & Miller, 2010).

Moreover, the test–retest reliability of fMRI measures probably hinges on the specific task demands and stimuli examined. For example, Sauder, Hajicak, Angstadt, and Phan (2013) reported that the reliability of fMRI measures of amygdala activation was adequate in response to fearful faces (ICCs ranged from 0.32 to 0.43), but inadequate in response to angry faces (ICCs ranged from 0.24 to 0.11). Because in classical test theory, validity is limited by the square root of reliability (Meehl, 1986), these findings suggest that the construct validity of some fMRI measures may be considerably lower than commonly assumed (Vul et al., 2009). In contrast, structural MRI measures appear to have considerably higher test–retest reliability (Kendler & Neale, 2010). For example, the stability of measures of cortical thickness as assessed by structural MRI is on the order of r = .95 (Wonderlick et al., 2009).

Making matters more complicated, the fMRI research center at which the study is conducted appears to account for about 8 percent of the variance in fMRI blood oxygen-level dependent
(BOLD) signal results, suggesting that the laboratory itself is a potential source of error in analyses (Costafreda et al., 2007). Another investigation revealed that the median ICC of fMRI findings across different imaging centers that contained identical hardware setups was only 0.22 (Friedman et al., 2008; see also Bennett & Miller, 2010).

All of these limitations may hamper the replicability of functional neuroimaging findings unless explicitly addressed. Adding to these replicability concerns are findings that the average statistical power of functioning brain imaging studies is only about 8%, which is considerably lower than in most other domains of psychological and psychiatric research (Button et al., 2013). Low power not only increases the chances of Type II errors (false negative results), but also boosts the likelihood of detecting findings associated with inflated effect sizes, a phenomenon known as the “winner’s curse” (Button and Munafò, in press). Although meta-analyses may partly address concerns regarding low-powered neuroimaging studies, they are not an adequate substitute for conducting adequately powered investigations. Indeed, to the extent that low-powered neuroimaging studies are susceptible to the winner’s curse, the effect sizes derived from meta-analyses of these studies will tend to be upwardly biased.

Clearly, these considerations impart an oft-neglected lesson: the psychometric rigor of laboratory indicators, including neuroimaging data, must be demonstrated rather than assumed. Two additional bodies of literature, both on the convergent validity of T data (Block, 1977), should suffice to drive this point home. First, although many authors appear to regard neuropsychological measures of prefrontal functioning (e.g., Wisconsin Card Sorting Task, Stroop Color-Naming Task, Tower of Hanoi) as largely fungible, the correlations among these measures are modest (typically below $r = .25$) and frequently statistically nonsignificant (Miyake et al., 2000). Nevertheless, these intercorrelations are almost always positive, suggesting that aggregating these measures (e.g., by combining them into a latent variable) may help to circumvent the difficulties introduced by the situational specificity of diverse prefrontal tasks. Second, data on even superficially similar delay of gratification (DOG) tasks demonstrates that these measures often display surprisingly low correlations. For example, Block (1977) showed that in a sample of children, two DOG tasks, one in which participants were informed that they could keep an attractively wrapped gift if they could resist opening it, and another in which participants were informed that they could eat additional candy if they could resist eating candy in front of them, correlated only $r = .13$. Similarly, a meta-analysis by Duckworth and Kern (2011) of 10 DOG tasks ($n = 523$) revealed a mean intercorrelation of only $r = .21$. As with neuropsychological measures of prefrontal functioning, these findings should remind us that even laboratory measures that appear to assess the same construct routinely yield low or modest correlations (see also Shilling, Chetrywandy, & Rabbit, 2002, for evidence of low correlations among different variations of the Stroop color-naming interference paradigm).

**Recommendations**

Researchers should be certain to bear in mind measurement error, stemming largely from situational uniqueness, when incorporating laboratory indicators of biological systems into their studies. In particular, whenever possible they should strive to administer multiple laboratory indicators of targeted constructs. Aggregating several laboratory measures into composite observed variables or into latent variables, ideally along with data drawn from different levels of analysis (e.g., self-report, interview data) should help to address the problems posed by measurement error (see Patrick et al., 2013, for an excellent example). With respect to neuroimaging data, boosting the number of fMRI runs can often increase reliability substantially (see Bennett & Miller, 2010, for a discussion of this and additional recommendations). Increasing sample size, ideally by capitalizing on collaborations across neuroimaging laboratories, can boost statistical power and diminish the odds of both Type II errors and the winner’s curse (Button et al., 2013).

**Challenge #3: biological and psychometric limitations of endophenotypes**

One of RDoC’s long-term goals is to replace, or at least supplement, the DSM sign and symptom approach with an approach based largely on endophenotypes. The concept of endophenotypes, which first appeared in entomology (John & Lewis, 1966) and was imported into psychology and psychiatry by Gottesman and Shields (1972; see also Gottesman & Gould, 2003; Waldman, 2005), refers to “internal phenotypes discoverable by a biochemical test or microscopic examination” (Gottesman & Shields, 1972, p. 19). These internal phenotypes could include biochemical markers, brain imaging findings, autonomic indicators, and performance on laboratory or neuropsychological tasks, among others. Some authors have explored the endophenotype concept to include measures of personality traits that ostensibly underpin psychopathology (Flint & Munafò, 2007). Endophenotypes are distinguishable from exophenotypes, which are the more traditional indicators of psychopathology, such as the signs and symptoms of disorders captured by DSM criteria.

The RDoC rationale for endophenotypes is compelling. The hope is that compared with exophenotypes, endophenotypes are closer to the gene end of the lengthy and circuitous gene-behavior pathway, and should therefore yield more fruitful clues regarding the correlates and causes of psychopathology (Kendler & Neale, 2010; Miller & Rockstroh, 2013). In the words of Gottesman and Gould (2003), endophenotypes hold out the promise of “simpler clues to genetic underpinnings than the disease itself” (p. 636). Presumably, compared with exophenotypes, endophenotypes should (a) be more heritable and (b) display a simpler genetic architecture (Gould & Gottesman, 2006; Waldman, 2005). Endophenotypes are also posited to be largely state-independent, and to manifest themselves regardless of whether the disorder is present (Gottesman & Gould, 2003).

From an RDoC perspective, endophenotypes could in principle provide “cleaner” and more construct-valid indicators of biological systems, such as negative and positive valence systems. Moreover, because many endophenotypes cut across traditional disorder categories, these phenotypes accord well with RDoC’s transdiagnostic approach and emphasis on identifying etiological processes that underlie multiple DSM conditions (Miller & Rockstroh, 2013).

Despite high initial expectations and provisional successes for certain conditions, such as schizophrenia (see Cannon & Keller, 2006; Miller & Rockstroh, 2013, for reviews), the promise offered by endophenotypes has often proven difficult to realize. In particular, it is not evident that the endophenotypes identified to date are necessarily more genetically informative than are traditional exophenotypes. Flint and Munafò (2007) examined this issue in meta-analyses of studies of catechol O-methyl transferase (COMT), an enzyme that metabolizes dopamine (among other neurotransmitters), and schizophrenia, a disorder long known to be associated in part with dopamine overactivity. Specifically, they tested whether the COMT genotype displayed higher effect sizes with presumed neuropsychological and psychophysiological endophenotypes of schizophrenia, such as performance on the Wisconsin Card Sorting Task, the N-Back Task, and P300 amplitude and latency, than with
DSM schizophrenia itself, Flint and Munafo found no evidence that 
the ostensible endophenotypes were more highly related to the 
COMT genotype than was schizophrenia. As the authors acknowled-
ged, their findings may be limited to the COMT genotype and 
should not be generalized to other genotypes, let alone to other 
disorders. At the same time, their findings suggest that in-
vestigators should not assume that candidate endophenotypes will 
necessarily yield higher effect sizes than do exophenotypes in ge-
netic studies (but see Tan, Callicott, & Weinberger, 2008, for a more 
sanguine view of the status of endophenotypes for schizophrenia). 
Moreover, many of the psychometric considerations already dis-
cussed for T data (Block, 1977; Epstein, 1980) apply here as well; 
endophenotypic markers based on single laboratory tasks may 
possess substantial amounts of situational uniqueness and there-
fore high levels of measurement error.

In addition, many authors have used the term endophenotype 
largely or entirely interchangeably with the term “intermediate 
phenotype” (e.g., Cannon, 2006; Lilienfeld et al., 2013). This sem-
antic equation presumes that the candidate endophenotype acts 
statistically as a mediator, in which the genetic effects “pass 
through” the endophenotypic marker on their way to the exophe-
notype (e.g., the measure of the disorder in question; Kendler & 
Neale, 2010). Indeed, one crucial assumption of the endopheno-
type concept is that “the effects of a particular gene on the endo-
phenotype are expressed — either in full or in part — through the endophe-
notype” (Waldman, 2005, p. 1533).

Nevertheless, the evidence that putative endophenotypes mediate the relation between genes and behavioral phenotypes 
appears to be sparse and inconsistent. For example, in a twin 
sample, Kendler, Neale, Kessler, Heath, and Eaves (1993) found that 
although neuroticism was associated with elevated rates of major 
depression, it did not mediate the association between genetic risk 
and major depression (see also Kendler & Neale, 2010). Waldman 
(2005) reported mixed findings concerning whether scores on the 
Trail Making Test mediate relations between dopamine genes and 
tention-deficit hyperactivity disorder (ADHD), with partial 
mediation for Trails A but no mediation for Trails B. Indeed, it is 
possible that certain putative endophenotypes lie causally down-
stream of the exophenotypes with which they are associated 
(Kendler & Neale, 2010). For example, P300 amplitude appears to 
be a valid endophenotype for a broad predisposition toward 
externalizing behavior and disinhibition (Patrick et al., 2009). 
Nevertheless, P300 amplitude is also exquisitely sensitive to 
attention (Polich, 2012). Hence, it is conceivable that this marker is 
a consequence, not an antecedent, of the attentional and motiva-
tional deficits that characterize externalizing disorders, such 
conduct disorder, antisocial personality disorder, and substance use 
disorder, all of which display high levels of co-occurrence and 
covariation with ADHD (Lilienfeld & Waldman, 1990; Torgersen, 
Gjervan, & Rasmussen, 2006).

**Recommendations**

The RDoC endophenotype approach has considerable potential, 
and should certainly be vigorously pursued. Nevertheless, in the 
absence of clear-cut evidence, researchers should not presume that 
potential endophenotypes are more heritable or genetically 
simpler than are traditional exophenotypes. Nor should they pre-
sume that potential endophenotypes necessarily display higher 
effect sizes than do exophenotypes in their relation with genetic 
loci (MacDonald & Krueger, 2013). Investigators should also test 
mediational models to demonstrate that endophenotypes function 
as intermediate phenotypes, as implied by the endophenotype 
concept (Kendler & Neale, 2010; Waldman, 2005). As noted earlier 
in the case of laboratory measures, which may themselves be 
endophenotypes in certain cases, developing composite or latent 
variables derived from multiple correlated endophenotypic 
markers may help to reduce measurement error and in turn boost 
construct validity.

**Challenge #4: distinguishing biological predispositions from their 
behavioral manifestations**

One conceptual quandary that appears to have received little 
explicit discussion in the RDoC literature is the distinction between 
biological predispositions to psychopathology and the behavioral 
manifestations of these predispositions. In this respect, the 
distinction between basic tendencies and characteristic adaptations 
in the personality trait literature provides a useful organizing 
framework (McCrae & Costa, 1995; see also Hankness & Lilienfeld, 
1997; McAdams & Pals, 2006). Basic tendencies are underlying 
personality traits, such as positive emotionality and constraint/ 
disinhibition (see Tellegen & Waller, 2008), whereas characteristic 
adaptations are the behavioral expressions of these traits (see also 
Cantor’s, 1996, distinction between the “having” and “doing” as-
pects of personality traits).

The key assumption of this differentiation is that individuals do 
not merely develop adaptations to their external environments; 
they also develop adaptations to their own traits. Some of these 
adaptations may be healthy for the individual, whereas others may 
be unhealthy. For instance, an individual with high levels of nega-
tive emotionality may manifest this predisposition in an anxiety 
disorder; alternatively, she may manifest it in artistic productivity, 
which is associated with a disposition toward negative emotion-
ality (Sheldon, 1994). As another example, data show that the mean 
sensation seeking scores of prisoners are essentially indistin-
guishable from those of firefighters (Zuckerman, 1994). It is plau-
sible, if not likely, that the trait of sensation seeking can be 
manifested in either (a) socially and personally destructive outlets 
(e.g., crime, substance abuse) or (b) socially and personally 
constructive outlets (e.g., firefighting, law enforcement) depending 
on psychosocial factors and moderating individual differences, such 
as impulse control and vocational talents/interests (Hankness & 
Lilienfeld, 1997).

Similarly, the concept of multifinality in the developmental 
psychopathology literature reminds us that early patterns of 
behavioral adjustment can often eventuate in a variety of different 
long-term outcomes (Cicchetti & Rogosch, 1996; see Franklin et al., 
2014, for a thoughtful discussion of the implications of multifinality 
and equifinality for RDoC). For example, a large proportion and 
perhaps a majority of such adolescents with conduct disorder de-
sist from antisocial behavior in adulthood (Byrd, Loeb, & Pardini, 
2012; Lahey et al., 1995). These differences in outcome probably 
reflect in part the heterogeneity of conduct disorder, as only a mi-
nority of individuals with this diagnosis (e.g., those exhibiting high 
levels of callous and unemotional traits; Byrd et al., 2012) appear to 
be at especially pronounced risk for later antisocial and criminal 
behavior. Nevertheless, some of these differences in outcome are 
likely to also reflect differential exposure to risk or protective in-
fluences during the course of development. For example, children 
with oppositional defiant disorder, a condition that overlaps sub-
stantially with conduct disorder, are more likely to persist in their 
difficult behavior over a four-year interval if their caregivers engage 
in negative parenting practices (August, Realmuto, Joyce, & 
Heckner, 1999). It is not known, however, whether the apparent 
influence of parenting on such persistence is purely environmental 
or instead reflects passive or evocative gene—environment corre-
lation (see Moffitt, 2005).

Data on discordant monozygotic (MZ) twins among individuals 
with major mental disorders, including schizophrenia, bipolar 
disorder, and major depression, afford another potential illustration
of multifinality. For example, for reasons that remain poorly un-
derstood, between 35% and 59% of the co-twins (Cardno & Gottesman, 2000) of MZ twins with schizophrenia do not meet
diagnostic criteria for the disorder. Some of this discordance may
reflect epigenetic differences within MZ pairs, in turn perhaps
stemming partly from differential twin exposure to environmental
influences (Dempster et al., 2011; Kato, Iwamoto, Kakuchi,
Kuratomi, & Okazaki, 2005; Petronis et al., 2003). It is certainly
possible that such epigenetic differences, if replicable, could
be accommodated within an RDoC framework. Moreover, at least
some of the differences between MZ twin pairs discordant for
schizophrenia may be manifested in biological measures, such as
hypofrontality and diminished cerebral volume (Noga, Aylward,
Barta, & Pearson, 1995; Weinberger, Berman, Sudath, & Torrey,
1992). At the same time, the literature on MZ twin discordance
in schizophrenia and other mental disorders raises the possibility that
individuals with similar biological risk factors, such as those
reflecting a predisposition to schizophrenia, may display this risk in
substantially different exophenotypes.

In the RDoC context, the allied principles of (a) basic tendencies
versus characteristic adaptations and (b) multifinality remind us
that individuals with overlapping or identical biological pre-
dispositions toward psychopathology may ultimately manifest
these predispositions in markedly different ways, in part as a
consequence of developmental and psychosocial factors. Hence,
although RDoC may be a valuable starting point for a new taxon-
omy of psychopathology, it may necessarily be incomplete, as it
may in many cases be unable to distinguish physiological risk fac-
tors for psychopathology from psychopathology itself (see also
Wakefield, 2014).

Recommendations

Researchers operating within the RDoC framework should bear
in mind that this framework may be best suited for identifying
biological predispositions to psychopathology, and that these pre-
dispositions may be expressed in numerous different outcomes,
only some of which may be pathological. As a consequence, they
may wish to avoid an exclusive focus on endophenotypes, which
are traditionally regarded as largely state-independent and as in-
dependent of the disorder’s presence or absence (Gottesman &
Gould, 2003). By relying solely on trait indicators, researchers
may be limited to studying the biological predispositions to disor-
ders rather than to the behavioral expressions of these pre-
dispositions. As a consequence, they may want to supplement their
assessment of endophenotypes with measures of more state-
dependent indicators, such as hypofrontality, which might be more
sensitive to the presence of absence of psychopathology.

More broadly, psychologists and psychiatrists must take seri-
ously the possibility that an RDoC framework may be insufficient
for the classification of psychopathology. Instead, a complete clas-
sification system may need to be twofold. It may need to include
indicators of both (a) biological predispositions and (b) the signs
and symptoms of mental disorders, which can reveal whether these
predispositions have been realized in maladaptive behavior. In this
respect, it may be more realistic to expect RDoC to supplement
rather than supplant systems that include a detailed assessment of
the signs and symptoms of disorder (see also Fulford, 2014; Insel
et al., 2010).

Concluding thoughts

RDoC is a calculated gamble. Given that incremental efforts to
improve the DSM by nibbling around its edges have not translated
into measurable declines in the morbidity and mortality of most
serious psychiatric disorders (Insel, 2009), NIMH’s substantial
investment of time and money in a rival paradigm would appear to
be worth the risk. Even if RDoC is not an unqualified success, the
knowledge yielded by the endeavor may contribute to valuable
insights regarding the classification and etiology of psychopath-
ology. In the unlikely event that RDoC is a wholesale failure, even
this information could be helpful, as it might suggest that the recent
movement to tether psychiatric classification and treatment pri-
marily to a biomedical paradigm is misguided (see Deacon, 2013;
Satel & Lilienfeld, 2013, for discussions) or at least premature
given the our present state of neuroscientific knowledge.

Nevertheless, the admittedly selective analysis presented here
raises the possibility that the RDoC initiative may be fraught with
largely unappreciated impediments. To the extent that in-
vestigators working within the RDoC framework take heed of these
challenges, the likelihood of RDoC’s ultimate success should be
maximized. In particular, RDoC researchers should be careful to
steer clear of the premature optimism that has often been associ-
ated with previous efforts to identify markers of biological systems
relevant to mental disorders.

For example, on the cusp of the brain imaging revolution in
psychiatry three decades ago, prominent psychiatrist and editor of
the American Journal of Psychiatry Nancy Andreasen (1984), wrote
that “as they improve and become more accurate, imaging
techniques and other laboratory tests for mental illness will
become part of standard medical practice during the coming years,
thereby improving the precision of diagnosis” (p. 260; see also
Pardes, 1986). Yet, when DSM-5 (American Psychiatric Associa-
tion, 2013) was released last year, nary a single neurobiological marker
was to be found anywhere in the manual’s diagnostic criteria.
Furthermore, the only laboratory markers in the manual were
incorporated for a subset of sleep disorders. To some extent, the
continued failure of neural indicators to assist in psychiatric diag-
nosis surely reflects the fact that many DSM categories themselves
are markedly heterogeneous (First, 2014).

Nevertheless, the glaring discrepancy between past hope and
present reality should give us pause (Satel & Lilienfeld, 2013). The
current inability of biological and other laboratory markers to
inform the classification of psychopathology probably also stems
from the often neglected methodological limitations of these in-
dicators, many of which have not satisfied the fundamental psy-
chometric standards demanded of other measures in psychology
and psychiatry (see Epstein, 1980). By attending thoughtfully to
issues of measurement error, ideally by (a) constructing better in-
dicators and (b) capitalizing on the power of aggregation across
indicators, RDoC investigators can develop composites or latent
variables that possess higher construct validity for their targeted
biological systems.

A half century ago, Shakow (1965), in many ways the founder of
contemporary clinical psychology, wrote that “psychology is
immodest” (p. 353). By that, he meant our field’s regrettable habit
of advancing claims that go well beyond the extant data. In the field
of psychiatry, others have more recently argued that DSM-5’s fatal
flaw was its hubris, namely, its eagerness to introduce changes in
diagnostic criteria in the absence of adequate evidence (Frances,

RDoC must avoid the same missteps (Frances, 2014). For RDoC
to succeed, it will need to proceed with humility and with full
recognition of the hard–won lessons of the past. From the vantage-
point of this essay, the four most valuable caveats to bear in mind
moving forward are that (a) the biological level of analysis, although
essential, will not be sufficient to understand psychopa-
therapy (Ilardi, Rand, & Karwoski, 2007; Lilienfeld, 2007; Miller,
2010), (b) the level of analysis of constructs (e.g., biological)
should not be confused with the level of analysis of indicators of
these constructs, (c) the psychometric properties of biological and
other laboratory indicators must be demonstrated empirically, not merely presumed, and (d) a system of biological predispositions toward psychopathology is not a classification system of psychopathology, although it can inform such a system. If, but only if, RDoC absorbs the valuable insights imparted by decades of research in psychometric, social, personality, cultural, and developmental psychology, it may begin to deliver on its ambitious promises.

References


