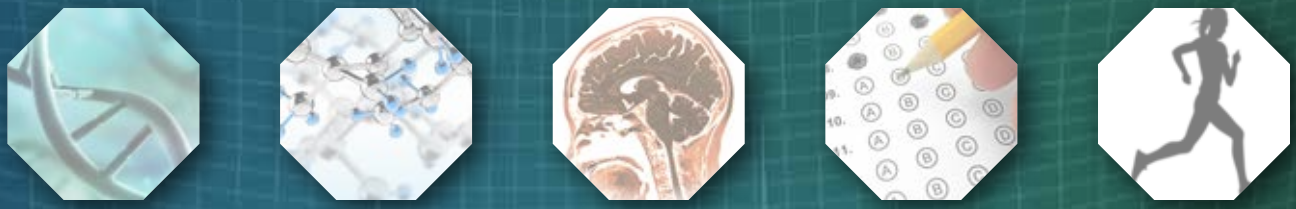
NATIONAL INSTITUTE
OF MENTAL HEALTH



Behavioral Assessment Methods for RDoC Constructs

August 2016

A Report by the National Advisory Mental Health Council Workgroup on Tasks and Measures for Research Domain Criteria (RDoC)

Department of Health and Human Services
Public Health Service
National Institutes of Health
National Institute of Mental Health

TABLE OF CONTENTS

SECTION I: EXECUTIVE SUMMARY

Introduction

The National Institute of Mental Health (NIMH) launched the Research Domain Criteria (RDoC) in 2009 in response to the 2008 NIMH Strategic Plan's call for new ways of classifying mental illnesses that are based on dimensions of observable behavioral and neurobiological measures. RDoC is a research framework designed to integrate many levels of information (from genomics to self-report) to better understand the basic dimensions of functioning underlying the full range of human behavior, from normal to abnormal. NIMH envisions that the RDoC initiative will determine how a classification approach based on biology, behavior, and context can be useful for mental disorders, thus informing diagnostic systems of the future.

Since its inception, RDoC has progressed as a significant effort for the Institute, impacting basic, translational, and services/intervention research priorities. Initially, a series of collaborative workshops was held in order to summarize the state of the knowledge related to five main "domains" and define associated constructs for each (see Appendix A). The current RDoC framework consists of a matrix in which the rows represent specified functional Constructs, concepts summarizing data about a specified functional dimension of behavior, characterized in aggregate by the genes, molecules, circuits, etc., which implement it. Constructs are in turn grouped into higher-level Domains of functioning, reflecting contemporary knowledge about major systems of cognition, motivation, and social behavior. In its present form, there are five Domains in the RDoC matrix: Negative Valence Systems, Positive Valence Systems, Cognitive Systems, Systems for Social Processes, and Arousal/Regulatory Systems. The matrix columns specify Units of Analysis used to study the Constructs, and include genes, molecules, cells, circuits, physiology, behavior, and self-reports. The matrix also has a separate column to specify well-validated paradigms used in studying each Construct. These paradigms may be relevant for more than one unit of analysis and rather than list them in separate columns, they are included under the Paradigms heading. In the body of the matrix are specific elements which are empirically associated with the construct and are grouped under the appropriate unit of analysis.

The RDoC matrix provides one framework for organizing NIMH research efforts, freeing scientists from traditional categories that are often heterogeneous and overlapping. RDoC aims to support research that considers mental illnesses in terms of fundamental behavioral-neural systems (e.g., fear or working memory) rather than traditional diagnostic categories. The long-term goal is to develop a scientific base that can inform future neuroscience-based diagnostic systems for mental illnesses. To generate a systematic RDoC database for this purpose, it is important to develop a set of paradigms and measures that are generally accepted by the field and which can facilitate comparisons across studies and sharing of data. However, if NIMH prematurely establishes for a battery of affective, behavioral, and cognitive tasks for use in RDoC research, it runs the risk of hampering future methodological innovation and revisions to the RDoC constructs, which would have deleterious effects on the long-term development of RDoC. A reasonable compromise is to establish a set of standardized paradigms and measures

which are appropriate for assessing RDoC constructs, but which are not required to be used in RDoC research. Such a list would offer the field some standardization that can foster data sharing through the RDoC Database (RDoCdb), but would require regular revision in order to incorporate new developments and findings.

To initiate the development of standardized paradigms and measures, NIMH's RDoC Unit proposed the concept clearance, *First Generation Research Domain Criteria (RDoC) Measurement Elements,* to the National Advisory Mental Health Council (NAMHC; see Appendix B). The Council approved this concept at its May 29, 2015 meeting. The aim of this initiative was to support the identification of two to four paradigms and/or measures that would be optimal for each RDoC construct. These measures would provide researchers a choice among a group of vetted elements, while still maintaining a degree of standardization. Identifying constructs for which no appropriate measures exist helps to identify areas in need of further assessment development.

The Workgroup on Tasks and Measures for RDoC

During the February 4, 2016 NAMHC meeting, NIMH Acting Director Bruce Cuthbert, Ph.D., announced the formation of the Workgroup to implement the *First Generation RDoC Measurement Elements* concept. A group of 34 researchers from 34 unique institutions was established (see Appendix C for a roster), with each participant agreeing to participate in one domain-specific subgroup. A leader was assigned for each domain subgroup, and took on the responsibility of leading the discussions and helping to assemble and coordinate the domain subgroup's final recommendations. The National Advisory Mental Health Council Workgroup on Tasks and Measures for Research Domain Criteria convened an in-person meeting on April 5 and 6, 2016 (see Appendix D for the meeting agenda) at the Neuroscience Center in Rockville, Maryland. Deanna Barch, Ph.D., Professor at Washington University and Maria Oquendo, M.D., Professor at Columbia University, co-chaired the Workgroup.

NIMH's Request for Information

In preparation for the workgroup meeting, NIMH published a request for information (RFI) titled "Building a Set of Recommended Tasks and Measures for the RDoC Matrix" on March 25, 2016, to seek input from the field. Responses to the RFI were due April 22, 2016. Through the RFI, NIMH gathered information about existing tasks and measurement tools that were recommended for inclusion in the RDoC matrix, as well as general suggestions about the most important criteria for consideration in selecting candidate tests.

As of May 10, 2016, NIMH received 60 responses. Of these, a subset of 42 were classified as relevant and on topic. Seven of these suggested general criteria to consider when selecting a task. The remaining responses included recommendations for specific tasks, across all five domains.

Workgroup Charge

The charge to the Workgroup was to recommend a set of two to four tasks for each construct that meet all or many of the following criteria. These criteria were developed based on discussions among the RDoC workgroup members prior to the start of the meeting, and modified through information gained from the RFI and from discussions at the start of the in-person Council workgroup meeting.

- How strong is the evidence that the task provides a valid measure of the RDoC construct?
- How good is the evidence about the psychometric characteristics of the task (e.g., internal reliability, test-retest reliability, floor and ceiling effects, practice effects, availability of alternate forms, and longitudinal stability)?
- Is the task free from floor/ceiling effects so that it can be used across individuals with the full range of performance/impairment on the tasks?
- Are parameters for administering the task (e.g., number of trials, stimulus characteristics, and primary dependent measure) standardized on an empirical basis?
- To what extent is the task (or different versions of the task) suitable for use across laboratory-based studies, clinical trials (as a measure of target engagement or clinical outcome), and/or high-throughput screening settings? Is the task suitable for use in human subjects in a variety of laboratory environments? Is the task feasible for administration across sites? What work is needed to make the task ready for use in clinical trials?
- Can the task be used (or adapted for use) with children and other special populations? Can it be used across different cultural settings?
- Can the task be used as a stand-alone behavioral task?
- Are adequate normative data available across age, gender, education, ethnicity, and socioeconomic status?
- Is the task widely used currently or has its use been limited to a few research groups?
- Is the task sensitive to within-person change?
- Are the relationships between task performance and clinical feature(s) known?
- Is the task freely distributed (i.e., not copyrighted)?
- Does the task assess multiple RDoC constructs or is it specific to just one? If it assesses multiple constructs, does it allow for unambiguous conclusions about the targeted construct?

During the domain-specific breakout sessions, the workgroup members were asked to rate each proposed task or measurement tool on each criterion using a scale of 1 to 5 (1 = no evidence, 3 = some evidence, 5 = strong evidence), in order to facilitate direct comparisons among task characteristics. Other proposed task characteristics considered important when evaluating a task, but not required for behavioral measures of RDoC constructs, are:

- Can the task be used with methods to interrogate brain circuitry (e.g., functional magnetic resonance imaging [fMRI] and EEG)?
- Can the task (or its analog) be used in animals? Is an animal version available?
- Are the relationships between task performance and neural signal(s) known?

In discussing the Workgroup's charge and criteria for task nomination, the members of the workgroup developed the following suggestions:

- Do not spend time at the workgroup meeting revisiting the organization of the matrix or the definitions of the constructs. Suggestions for changes are welcome but the focus of the discussions should be on measures for the constructs as they are currently defined.
- Wherever possible, the measures should allow for behavioral assessment, as opposed to focused solely on biological signals (e.g., neuroimaging). However, it was recognized that some constructs (e.g., sleep cycles) cannot be measured behaviorally. NIMH will obtain recommendations for tasks and measures related to levels of analysis—including electrophysiology and neuroimaging measures—in future meetings.
- When choosing among measures, a task that relates to clinical features (particularly functional status) is preferred.
- Regarding the use or adaptation of tasks for children and other special populations, workgroup members should consider whether the test is sensitive to normative developmental change. With a task that both children and adults can perform, it would be helpful to be able to determine whether the groups are using the same or different strategies.
- Workgroup members should consider that some measures are influenced by culture.
- Workgroup members should consider the acceptability of tasks to subjects. Some might be too difficult and perceived difficulty may vary across population groups.
- The new web design of the matrix allows for the addition of information (e.g., references to publications) about elements. Therefore, workgroup members should identify information about tasks that can be added to the matrix.
- When nominating a task, the workgroup should note where possible:

  - the particular psychometric properties of the task or paradigm (where information is available) and the subpopulations that have been tested;
  - whether the task measures a state or trait;
  - the appropriate use of the task (e.g., whether it is suitable for longitudinal research versus single administration);
  - whether the parameters for administering a task (e.g., number of trials, stimulus characteristics, and primary dependent measures) have been standardized based on empirical evidence;
  - variation in the parameters needed to obtain the desired level of sensitivity across populations; and
  - the settings in which the task can be used (e.g., laboratory or clinical).

The tasks currently listed in the RDoC matrix provided a starting point for the workgroup's deliberations. Members of the workgroup were also encouraged to identify other tasks which may be well-suited for specific constructs and to identify constructs for which new tasks are needed. The workgroup was informed that a successful report would provide (1) a list of currently-available tasks and measures that are recommended for inclusion in the RDoC battery, (2) a list of tasks that could be appropriate for inclusion but are in need of further

optimization and a summary of the work needed in order to optimize them, and (3) a list of constructs for which no appropriate tasks are available. Workgroup members were also asked to provide a list of paradigms that were considered for inclusion but not recommended, including the rationale for exclusion.

The intent of these recommendations from the workgroup is not to be overly prescriptive; the goal is to facilitate use of common data elements where feasible. The list of recommended tasks will be dynamic, as researchers in the field will be able to make the case for other tasks or measures that also meet the criteria.

Summary of Recommendations and Conclusions

General Issues

In their discussions of tasks and measures, all domain subgroups encountered a particular challenge: **the absence of psychometric data.** For many of the recommended tasks there are no normative data. The field would benefit from additional data, and further analysis, in order to understand the basic psychometric properties of popular tasks in current use. Similarly, many of the tasks do not have empirically derived administration parameters, and lack standardization across sites. Further optimization and standardization to ensure that all labs using a certain task are measuring the same phenomenon would be useful.

It is also noted that a number of the domain subgroups focused on behavioral measures and did not consider self-report measures, in large part because of time constraints at the workgroup meeting. The lack of self-report recommendations should not be interpreted to mean that the workgroup considered these to be invalid or not recommended, but perhaps should be the focus of a future meeting.

Another issue that came up in many domain subgroups was the question of how to address regulatory processes, including emotion regulation. Emotion regulation is currently considered to be an implicit component of any pertinent construct; for example, control of fear behavior or control of impulsive behavior. Thus, emotion regulation was not originally defined as a distinct construct in any domain. However, the domain subgroups suggested that it may need a more explicit role in the matrix. As such, the domain subgroups suggested that more focused discussion of methods for assessing this critical concept were needed.

Lastly, many domain subgroups noted that the organization of domains, definitions of the constructs, or overall scope/coverage of the field would benefit from updating. Some domain subgroups (i.e., Positive Valence) made specific suggestions as to how to change the Domain, whereas other domain subgroups (i.e., Negative Valence) simply noted that the organization and definitions were difficult, but worked within the guidelines to recommend tasks for the existing matrix. It is recommended, however, that the definitions and organizations of the constructs be evaluated in a future meeting. During the May 26, 2016 NAMHC meeting, NIMH Acting Director Bruce Cuthbert, Ph.D., announced the formation of a new Workgroup on Revisions to the RDoC Matrix. The charge to this group will be to advise the NIMH on

modifications to the RDoC matrix, including addition of new domains and constructs. Many of the recommendations about domain organization made by the workgroup on Tasks and Measures for RDoC will be passed along to this newly formed workgroup for their discussion.

Domain-specific Task Recommendations

Each of the domain subgroups provided an extensive final report. Here we provide executive summaries. The full reports can be found in Section II: Domain Specific Reports.

Negative Valence Systems (See Section II for full report)

The Negative Valence Systems subgroup noted difficulties with the way the domain was defined and organized. They suggest that several of the construct definitions do not lend themselves to a laboratory measurement model that would elicit the individual differences of interest. For example, many of their recommendations for Sustained Threat and Loss actually induce analogs for the affective state, or measure downstream consequences, and do not tap the defined construct directly. Additionally, they suggest that the domain is lacking in coverage across the topic area, and should more explicitly dovetail with Positive Valence Systems, as there is a great deal of overlap in the tasks and measures that could be used. They support the addition of constructs of “emotional lability”, “pain”, and “affective decision making”.

Negative Valence Systems Recommended Task Paradigms

Construct/Sub-construct	Task	Key references
1. Acute Threat		
	Trier Social Stress Test	Kirschbaum et al. 1993 Allen et al. 2014
	Behavioral Approach Test	none listed
	Cold Pressor Test	Edelson et al. 1986 Velasco et al. 1997 Rolke et al. 2006
	CO^2 Challenge	none listed
	Stranger Tests	Buss et al. 2003 Pfeifer et al. 2002
	Fear Conditioning Tasks	Norrholm et al. 2008 Zeidan et al. 2012
2. Potential Threat		
	No Shock, Predictable Shock, Unpredictable Shock (NPU Threat Task)	Schmitz et al. 2012
3. Sustained Threat		
	None (see full report for discussion of why none were recommended)	
4. Loss *(analog of response to loss*		
	Sadness eliciting film clips *(but only with w/immersion instructions and facial expression or mood ratings as dependent variables of interest)*	Samson et al. 2015 Joormann et al. 2007
5. Frustrative Nonreward		
	Points Subtraction Aggression Paradigm (PSAP)	Cherek, 1981 Geniole et al. 2016
	Laboratory Temperament Assessment Battery tasks of Box Empty and Transparent Box	Gagne et al. 2011

Negative Valence Systems Recommended Self Report Measures

Construct/Sub-construct	Task	Key references
1. Acute Threat		
	Subjective Unit of Discomfort Score (SUDS)	Wolpe, 1990 Kaplan et al. 1995
	Fear Survey Schedule	Wolpe & Lang, 1977
2. Potential Threat		
	Intolerance of Uncertainty Scale (12 item version)	Carleton et al. 2007
	Behavioral Inhibition Scale (BIS)	Carver & White, 1994
	Fear of Negative Evaluation Scale	Watson & Friend, 1969
	Anxiety Sensitivity Index	Taylor et al. 2007
	Life Events and Difficulties Schedule (LEDS)	Brown & Harris, 1978
3. Sustained Threat		
	Youth Life Stress Interview	Rudolph & Flynn, 2007
	Childhood Trauma Questionnaire	Bernstein & Fink, 1998
	LEDS difficulties	Brown & Harris, 1978
	Traumatic Events Screening Inventory (TESI)	Ippen et al. 2002
	Risky Families	Taylor et al. 2004
	Stress and Adversity Inventory (STRAIN)	Slavich & Epel, 2010
4. Loss *(analog of response to loss)*		
	LEDS (social experience of loss and potential threat)	Brown & Harris, 1978
	STRAIN	Slavich & Epel, 2010
5. Frustrative Nonreward		
	Frustrative Nonreward Responsiveness Subscale	Wright et al. 2009
	Questionnaire of Daily Frustrations	Baars et al. 2011

Positive Valence Systems (See Section II for full report)

The list of suggested tasks and measures for this domain reflects a slight regrouping and renaming of the constructs to more clearly match the existing empirical literature, which the Positive Valence Systems subgroup suggests reduces potential redundancies across the constructs, and isolates "purer" constructs. They propose 3 total Constructs; "Reward Responsiveness", "Reward Learning" and "Reward Valuation," each with 3 new sub-constructs. The domain subgroup discussed the fact that many tasks that were developed early and have been widely used often conflate multiple sub-constructs. Thus, many of these tasks might subsume different sub-constructs in the same task. New paradigms have less accumulated data but are more precise in differentiating sub-constructs. As with the other domains, the workgroup also noted that much more data are needed on psychometrics and norms for most if not all of the tasks. Additionally, the group suggested that the regulation of the Positive Valence Systems constructs (e.g., modulation of PVS constructs by homeostatic drives like hunger, sleep, thirst, sex) would involve processes that are better captured by Cognitive Systems and Arousal and Regulatory Systems, and so they did not include tasks that probed these regulatory processes in their deliberations.

Positive Valence Systems Recommended Task Paradigms

Construct/Sub-construct	Task	Key references
1. Reward Responsiveness		
1.1. Initial Response to Reward	Simple Guessing Task	Delgado et al. 2000 Carlson et al. 2011
1.2. Reward Anticipation	Monetary Incentive Delay Task	Knutson et al. 2000
1.3. Reward Satiation	Fixed-ratio Satiation Schedule	Sherman & Thomas 1968
2. Reward Learning		
2.1. Habit	Devaluation Task	Gillan et al. 2011
	Habit Task	McKim et al. 2016
	Habit Learning Task	Tricomi et al. 2009
2.2. Probabilistic and Reinforcement Learning	Probabilistic Reward Task	Pizzagalli et al. 2005
	Pavlovian Conditioning	O'Doherty et al. 2004
	Drifting double bandit	Daw et al. 2011
	Probabilistic Stimulus Selection Task	Frank et al. 2004
2.3. Reward Prediction Error	Rutledge Passive Lottery Task	Rutledge et al. 2010
	Drifting double bandit	Daw et al. 2011
3. Reward Valuation		
3.1. Reward (probability)	Probability Choice Task	Levy et al. 2010
	Willingness To Pay Task	Becker et al. 1963
3.2. Delay	Delayed Discounting Task	Kable & Glimcher 2007 Johnson & Bickel 2002 Green & Myerson 2004
3.3. Effort	Effort Expenditure for Reward Task	Treadway et al. 2009

Cognitive Systems (See Section II for full report)

The Cognitive Systems Domain subgroup group discussed the fact that many cognitive constructs overlap (for example, working memory and cognitive control), and that this is the nature of cognition and to some extent unavoidable. There is additional overlap between the Cognitive Systems Domain and other domains (for example, vigilance is an aspect of attention and also an index of arousal.) The domain subgroup also noted that some key cognitive constructs were not currently represented in the matrix, such as reasoning and inference. The domain subgroup has suggested an update of the Attention construct, in light of current work in cognitive neuroscience and suggests three subconstructs, "Controlled vs. Automatic Attention," "Capacity and Interference Control," and "Vigilance (Sustained Attention)." Another observation was that the construct "Language Behavior" was less well elaborated than other constructs. This domain subgroup felt, given the specialized nature of the field of linguistics and the interactions between linguistic and cognitive systems, that identifying subconstructs and paradigms from this construct would be best accomplished by a new subgroup with more expertise in the area.

Cognitive Systems Recommended Task Paradigms

Construct/Sub-construct	Task	Key references
1. Attention		
1.1. Overt/Covert	Spatial and non-spatial cuing tasks	Carter et al. 1992
	Attention Networks Task (ANTS)	Macleod et al. 2010
	Visual search paradigm	Gold et al. 2007
1.2. Capacity and Interference Control	Attentional blink during rapid serial visual presentation	Mathis et al. 2011
	Dual task paradigms	Nuechterlein et al. 2006
1.3. Vigilance	Tasks with 'catch' trials (change detection working memory, perceptual threshold effects)	Barch et al. 2011
	Mind-wandering tasks	Smallwood & Schooner, 2015
2. Perception		
2.1. Visual	Contrast-Contrast Task	Barch et al. 2011
	Jittered Orientation visual integration task (JOVI)	Silverstein et al. 2011
3. Declarative Memory		
	Relational and Item Specific Encoding Task (RISE)	Ragland et al. 2012
	Mnemonic Similarity Test	Bakker et al. 2008
4. Cognitive Control		
4.1. Goal Selection, Updating, Representation and Maintenance	Continuous Performance Tests (AX and DPX)	Lopez-Garcia et al. 2015
	Preparing to overcome prepotency task (POP)	Snitz et al. 2005
4.2 Response Selection, Response Inhibition/Suppression	Go/No-go tasks	Boucher et al. 2007
	Stop Signal Tasks	Luijten et al. 2014
4.3 Performance Monitoring	Flanker Task versions	None
	Simon Task versions	None

Construct/Sub-construct	Task	Key references
	Stroop Task versions	Kerns et al. 2004
5. Working Memory		
5.1. Active Maintenance	Match to Sample	Horwitz & Tagaments, 1999
	Sternberg tasks	Nelson et al. 2003
	Change Detection	Barch et al. 2011
	Continuous Performance Tests (AX and DPX)	Lopez-Garcia et al. 2015
5.2 Flexible Updating	NBack tasks	Jonides et al. 2008
	Self-ordered Pointing	Gillett, 2007
5.3 Limited Capacity	Change Detection	Barch et al. 2011
5.4 Interference Control	Nback tasks	Jonides et al. 2008
	Sternberg tasks	Nelson et al. 2003

Systems for Social Processes (See Section II for full report)

The Systems for Social Processes subgroup suggested some changes to the domain's organization, and suggested adding "Rejection Sensitivity" and "Social Motivation" as subconstructs under the Affiliation and Attachment construct. The group noted that the "Social Communication – Production of Facial Communication" subconstruct would benefit from further development regarding methods of eliciting emotions and measuring facial expressions. Beyond facial communication, there is a significant need to develop techniques and instruments that capture the dimensionality of functioning across the life span, as well as instruments that maximize ecological validity.

The domain subgroup strongly recommended eliminating the Strange Faces (separation-reunion) task, the Still Face, and the Ford Corollary Discharge paradigms from the list of paradigms currently listed in the RDoC matrix. The group identified significant problems with these tasks; however this does not mean they endorse all of the remaining tasks in the current matrix.

Systems for Social Processes Recommended Task Paradigms

Construct/Sub-construct	Task	Key references
1. Affiliation and Attachment		
1.1 Rejection Sensitivity	Cyberball	Hartgerink et al. 2015 Bolling, 2011
1.2 Social Motivation	One-armed Bandit Task	Lin et al. 2012
2. Social Communication		
2.1. Reception of Facial Communication	ER-40 – Penn Emotion Recognition Test	Erwin et al. 1992
	Gaze Cuing	Gross & Levenson, 2008
2.2. Production of Facial Communication	None	
2.3. Non-facial communication (merged reception and production)	TASIT 1	McDonald et al. 2003
3. Perception and Understanding of Self		
3.1. Agency	None	
3.2. Self Knowledge	Self-Referential Memory Paradigm	Kelley, et al. 2002
4. Perception and Understanding of Others		
4.1. Animacy Perception	Point Light Displays of Biological Motion	Bjornsdotter et al. 2016
4.2. Action Perception	How part of How/Why task	Spunt & Adolphs, 2014
4.3. Understanding Mental States	Hinting Task	Corcoran & Frith, 2003
	Reading the Mind in the Eyes	Vellante et al. 2013

Systems for Social Processes Recommended Self Report Measures

Construct/Sub-construct	Task	Key references
1. Affiliation and Attachment		
1.2 Social Motivation	Multidimensional Scale of Perceived Social Support	Zimet et al. 1988
2. Social Communication		
2.3. Non-facial communication (merged reception and production)	Social Responsiveness Scale	Constantino et al. 2003

Arousal and Regulatory Systems (See Section II for full report)

The Arousal and Regulatory Systems subgroup worked directly from constructs already defined in the RDoC matrix, and did not suggest any revisions or edits to those constructs. The group pointed out some issues with the general concept of "arousal," indicating that it is not well-defined in the matrix, and that the term generally cuts across many constructs in domains, including attention, motivation, and anxiety, among others. The group also noted that the constructs are subserved by a wide array of neurobiological processes and functions, which adds to the complexity of trying to disentangle arousal from other domains of the matrix.

They note that many of the measures that were considered do not have agreed upon standards for administration or analysis, and most need more normative data. There are several recommended measures and tasks that include both autonomic nervous system and the central nervous system. The group suggests that polysomnography, or sleep EEG, is a very useful and widely used tool for the Sleep-Wakefulness construct and a much better measure than home recordings, but acknowledge that it is time consuming and expensive. Lastly, the group notes that there are not many good self-report measures in this domain and development work on these may be of benefit to the field.

Arousal and Regulatory Systems Recommended Tasks

Construct/Sub-construct	Task	Key references
1. Arousal		
	Heart Rate Variability (HRV)*	Beauchaine et al. 2015
	Electrodermal Responding (EDR)*	Boucsein et al. 2012
	Pupillometry*	Beatty et al., 2000
	Cardiac Pre-ejection Period (PEP)*	Sherwood et al. 1990
	Psychomotor Vigilance Task†	Basner et al. 2011
2. Sleep-Wakefulness		
	Latency to persistent sleep (LPS), Wake time after sleep onset (WASO), Total sleep time (TST)‡	Iber et al. 2007
	Sleep Spindles‡	Iber et al. 2007
	Non-REM Sleep, Sleep EEG Slow Wave Activity‡	Dijk et al, 1993
	Multiple Sleep Latency Test (MSLT)‡	Littner et al. 2005
	Insomnia Severity Index§	Bastien, 2001
	Finger Tapping Motor Sequence Task (MST)	Karni et al. 1998
3. Circadian Rhythms		
	Dim Light Melatonin Onset (DLMO)	Burgess et al. 2015
	Longitudinal Actigraphy	Briscoe et al. 2014
	Morningness-Eveningness Questionnaire (MEQ)§	Horne and Ostberg, 1976
	Munich Chronotype Questionnaire§	Roenneberg et al, 2003

* Autonomic measure

† Cognitive measure

‡ All measured by polysomnography

§ Self-Report measure

Next Steps

Both the proceedings of this workshop and the advances in the field over the past several years suggest a number of important next steps in the RDoC Initiative. We outline them briefly here:

Critical Evaluation of Current RDoC Domains and Constructs: The field is learning a great deal about the types of domains and constructs included in RDoC as the pace of research on these constructs has evolved. The domain subgroup reports make it clear that changes are likely needed to some of the domains and constructs given new knowledge about their validity and their organization. Such changes are likely to make the RDoC framework more useful in terms of generating information about putative brain-behavior dimensions relevant to psychopathology. Thus, a new round of workshops to evaluate and instantiate these changes would be useful. As with workshops on other levels of analysis, should such workshops be envisioned, we would recommend an early start to gathering information from the field, ideally through the use of a more focused survey in lieu of an RFI approach, which though helpful, is more general.

Development of New RDoC Domains or Constructs: Results of the domain subgroups' work indicate some areas where new constructs or domains are needed, for example, emotion regulation. This critically important construct is not currently well captured in any existing RDoC domain, and further consideration could help determine if it would be beneficial to modify the current view of emotion regulation as implicit in relevant constructs (e.g., fear, reward-related activity, lack of cognitive control). It is essential to develop a process by which new domains or constructs could be proposed and the evidence for their validity systematically evaluated. Such a process would benefit from explicit consideration of recommended tasks and paradigms across different levels of analysis, similar to the process undertaken by the current workgroup to identify behavioral tasks and paradigms.

Related to both of these recommendations, a new Council workgroup that will advise NIMH regarding changes and updates to the RDoC matrix was established in May 2016 and will have its first meeting in September 2016.

Analogous Process for Other Levels of Analysis: This workgroup focused on measures with "behavioral" outputs, primarily due to the need to focus the evaluation efforts to meet time and practicality constraints. However, as noted in several domain subgroup reports, in some cases, a different level of analysis may either be the only way to measure a given construct, or may be a better way to measure that construct. As such, additional workgroups that go through a similar process with measures at other levels of analysis, such as neuroimaging measures and/or peripheral physiology, will be essential. Similarly, few domain subgroups had time to systematically evaluate self-report measures for many constructs, and a workgroup specifically focused on self-report would also be beneficial. Should such workshops be envisioned, we would recommend an early start to gathering information from the field, ideally through the use of a more focused survey in lieu of an RFI approach, which though helpful, is more general.

Developmental Considerations: The field is increasingly focused on early detection and identification. To accomplish this goal, research on RDoC related constructs needs to be conducted in children, including very young children. For example, the Negative Valence Systems subgroup was able to provide a strong integration of developmental considerations. It is highly likely that many promising paradigms validated in adults will only work effectively with children if they are either modified (simpler instructions or tasks, developmentally appropriate materials, etc.), or use a different approach to measurement (observational measures, etc.). Such developmental considerations will continue to lag behind if not specifically prioritized, either through focused workshops or through research with the specific goal of making developmentally appropriate modifications to paradigms useful in adult populations. Similar concerns may arise when extending RDoC related work into geriatric populations, where other types of lifespan appropriate task modifications may be needed.

Standardization and Psychometric Evaluation: A few of the recommended tasks described above have standardized versions with at least some data about their psychometric properties. However, every domain subgroup noted that even for many promising paradigms or classes of paradigms, little standardization of administration parameters exists and in many cases, little psychometric data exist. In order to achieve the common data elements goal, it will be crucial for there to be: (1) standardization with appropriate attention to potential variation needed as a function of population and (2) evaluation of the psychometric properties of these tasks. It is unlikely that common data elements will be adopted for many constructs until this work is done. Final measures that are widely and freely available on flexible and easy-to-use platforms will facilitate data sharing and integration, but they must be undergirded by this key groundwork.

In summary, development of the RDoC system will require focused attention to ensure that the domains and constructs remain informed by new evidence and are refined as more work is conducted. The key goal of identifying common data elements to facilitate data sharing and comparisons across laboratories will necessitate similar processes to the ones described here for different levels of analysis. Although incremental, such steps are critical to enhancing the quality of data available to address the underlying neurobiological mechanisms of behavior ranging from normal to abnormal.

SECTION II: DOMAIN SPECIFIC REPORTS

The following reports were generated by each domain subgroup, based on their discussions both at the meeting, and after the meeting was complete.

Negative Valence Systems Final Report

C. Emily Durbin, Ph.D., Ian H. Gotlib, Ph.D. Sheri L. Johnson, Ph.D., Mercedes Perez-Rodriguez, M.D., Ph.D., Stewart Shankman, Ph.D. (chair)

I. GENERAL COMMENTS

The NVS subgroup was charged with developing a list for the five constructs listed within NVS domain - (1) Acute Threat; (2) Potential Threat; (3) Sustained Threat; (4) Loss; and (5) Frustrative Nonreward. Given this charge and discussion at the outset of the meeting, the NVS group decided to work strictly from the constructs and existing definitions listed in the RDoC matrix (http://www.nimh.nih.gov/research-priorities/rdoc/constructs/rdoc-matrix.shtml) and not attempt to revise, add, or clarify the constructs in the matrix.

A first important issue that significantly guided the NVS subgroup's discussion was the fact that for several constructs, the committee charge proved to be difficult; the definitions of the constructs did not always lend themselves to a measurement model in which the eliciting contexts for the individual differences of interest could be recreated via specific laboratory or *in vivo* paradigms. For example, Sustained Threat (NVS construct #3) is defined as "An aversive emotional state caused by prolonged [i.e., weeks to months] exposure to internal and/or external condition(s), state(s), or stimuli that are adaptive to escape or avoid." We felt that there are no paradigms that could be used ethically to assess directly the effects of sustained threat in humans. We acknowledge that there are real-life situations that might be used as quasi-experimental paradigms to assess the effects of sustained threat in humans (e.g., combat exposure, natural disasters). In addition, while we could identify paradigms that assess 'downstream consequences' of sustained threat (e.g., attentional vigilance to emotional stimuli), the specificity of these consequences to sustained threat (as opposed to acute threat, potential threat, or threat in general) was not clear. The group had a similar difficulty with the construct of Loss. Loss is defined in the RDoC matrix as "a state of deprivation of a motivationally significant con-specific, object, or situation… and may include permanent or sustained loss of shelter, behavioral control, status, loved ones, or relationships." We felt that this specific affective state could not be induced through the use of laboratory/in vivo paradigms in humans– thus, we listed paradigms that induce analogs for this affective state as well as stressful life events interviews that probe past experiences of loss. It is important to highlight that this issue did not apply to our discussion of paradigms that assess Acute Threat, Potential Threat, and Frustrative Nonreward because there are well-established paradigms that assess each of these three constructs.

There was a second important issue that significantly guided the NVS subgroup's discussion. The overall workgroup was instructed specifically to identify only paradigms for which there were clear behavioral outputs. Thus, for example, paradigms that elicited only a neural response and no behavioral output (e.g., Hariri Hammer Task[1]) were not included. Although this parameter made sense given the broader aims of the workgroup, and the fact that future meeting would focus on additional levels of analysis, this had the effect of narrowing the types of paradigms that could be listed.

Finally, although the NVS domain subgroup was able to provide ratings for most of the 18 criteria for each proposed paradigm, the group was struck by the consistently low ratings for several of the criteria. For example, the field lacks normative data for many of the paradigms (criterion #8). In addition, while there are conventions in the field for how several of the paradigms should be administered (e.g., number of trials, duration, etc.), most of these parameters have not been empirically determined or assessed (criterion #4). The NVS subgroup felt that these are important areas for future research.

II. ORGANIZATION OF THE DOMAIN

1. Concrete suggestions for changes to constructs: additions and deletions.

   As we noted above, our group did not spend a great deal of time discussing the structure and organization of domains; however, we present below several specific suggestions that arose during our meeting.

   (A) The NVS and positive valence system (PVS) domains are closely related in that both refer to responses to motivationally salient stimuli, but of different valences. Paradigms that provide opportunities for observing behavioral profiles relevant to one domain often have conditions (or versions) that elicit evidence of individual differences in the other domain as well. Identifying areas of overlap and distinction between the NVS and PVS constructs, both conceptually in terms of psychological processes and methodologically in terms of best practices for establishing convergent and discriminant validity, should be a priority for future work.
   (B) Further consideration should be given to adding the following constructs to the negative valence domain: 1) emotional lability (and other aspects of the time course of affective responding such as affective chronometry); 2) pain, and; (3) affective decision making.

2. Rationale for recommended changes. The suggested constructs listed above are highly relevant for severe mental illness, have well-studied neural circuits, and are not represented in other domains of the RDoC.

III. RECOMMENDED TASKS

1. Paradigms (See Appendix NVS-I for ratings of task criteria for the following paradigm recommendations).

   ACUTE THREAT: 1) Trier Social Stress Test and similar social performance tasks[2,3]; 2) Behavioral Approach Test (e.g., fear & disgust stimuli); 3) Cold Pressor (and other pain tolerance tasks)[4-6]; 4) CO^2 Challenge[7,8]; 5) Stranger Tasks[9,10]; 6) Fear Conditioning Tasks (an important correlate of Acute Threat)[11,12]

   POTENTIAL THREAT: No Shock, Predictable Shock, Unpredictable Shock (NPU-Threat Task)[13]

SUSTAINED THREAT: None

LOSS (analog of response to loss): Sadness-eliciting film clips, but only with w/immersion instructions and facial expression or mood ratings as dependent variables of interest[14,15]

FRUSTRATIVE NONREWARD: 1) Points Subtraction Aggression Paradigm (PSAP)[16,17]; 2) Laboratory Temperament Assessment Battery tasks of Box Empty and Transparent Box[18]

2. Self-Report

ACUTE THREAT: manipulation check measures (e.g., SUDS[19,20]), trait or experience measures or feared stimulus identification measures (e.g., Fear Survey Schedule[21])

POTENTIAL THREAT- Intolerance of Uncertainty Scale (12 item version[22]), Behavioral Inhibition Scale (BIS)[23], Fear of Negative Evaluation Scale[24], Anxiety Sensitivity Index[25], Life Events and Difficulties Schedule (LEDS)[26]

SUSTAINED THREAT = Youth Life Stress Interview[27], Childhood Trauma Questionnaire[28], LEDS difficulties, TESI[29], Risky Families[30], Stress and Adversity Inventory (STRAIN)[31]

LOSS = LEDS (social experience of loss and potential threat)[26], STRAIN[31]

FRUSTRATIVE NONREWARD = Frustrative Nonreward Responsiveness Subscale[32]; Questionnaire of Daily Frustrations[33]

IV. TASKS THAT REQUIRE FURTHER EVALUATION

Construct: Acute Threat

Paradigm: IAPS pictures or Viewing of Emotion Inducing Films

This paradigm has the potential to measure acute threat, but the NVS subgroup felt that this would only be the case if the stimulus set was restricted to particularly threatening stimuli and not simply those that are more broadly 'negative' in valence. As an example, trauma-specific stimuli (e.g., helicopters, humvees) for veterans with trauma-related psychopathology would be an appropriate use of this paradigm to measure Acute Threat.

Construct: Sustained Threat (more accurately, *consequences* of experiencing sustained threat)

Paradigm: Dot-Probe Task (to assess vigilance or attentional capture), Exogenous Cuing Task (to assess inability to disengage from particular classes of stimuli), Facial Morphing Task (for detecting threat thresholds)

As discussed above, the constructs measured by these tasks do not directly measure individuals' response to a sustained threat, but rather, assess constructs that are consequences of having previously experienced sustained threat (at least given the

definition of Sustained Threat in the RDoC matrix). The NVS subgroup had an additional concern with the Dot Probe task. Despite its widespread use, several studies have raised questions about its reliability (e.g., Staugaard, 2009 - *Psychological Science Quarterly*, although see Price et al., 2015-*Psychological Assessment* for a recent report in which adequate reliability was obtained using novel methods). There are, however, multiple variants of the dot-probe task (e.g., supraliminal presentation, subliminal presentation, verbal vs. pictorial stimuli, etc.), and the psychometric properties of the different versions of the dot-probe are likely to differ. The dot-probe task did, however, achieve high ratings for several of the other criteria. This raised another issue concerning whether certain criteria should be weighted more heavily than others in determining whether a paradigm should be recommended for a specific RDoC construct. Criteria ratings for the Dot-Probe and Exogenous Cuing Tasks are provided in Appendix NVS-II.

V. TASKS THAT ARE NOT RECOMMENDED

The NVS subgroup decided to focus on exemplar tasks for each construct rather than discuss paradigms that are in use and then attempt to fit them to specific RDoC construct. Below are a list of paradigms that the group discussed but that were judged to not fit directly into the definitions of the five NVS RDoC constructs, and/or that may be better represented by other RDoC constructs.

1. *Explicit (but not implicit) emotion regulation paradigms* - including instructions of distance, suppress, accept, maintain. Excluded because these paradigms appear to address regulation, which is not one of the current Negative Valence RDOC constructs.
2. *Darkness in humans/light in rodents* (excluded due to little behavioral yield as a paradigm)
3. *Flanker Task to Assess Response to Errors* – Unclear whether it elicits a threat response (as per the definitions of the three RDoC threat constructs). Despite clearly being negative in valence, this paradigm may perhaps belong with measures of cognitive control.
4. *Loss learning/loss aversion* – Did not fit well with acute threat or brain's defensive motivation system more broadly. Perhaps is more appropriate within the cognitive system.
5. *Questionnaires assessing Symptom Dimensions (e.g., guilt, shame, bereavement)* These questionnaires were excluded as they tap outcomes of the RDoC constructs rather than inductions of the RDoC constructs
6. *Autobiographical memory probes* (for measuring loss and threat) – These paradigms can elicit sensations of loss, threat, etc. that are idiographic in nature. However, the NVS subgroup excluded them given the difficulty of standardization for strictly behavioral outcome measures. These paradigms may be useful, however, for eliciting broader negative affective states.

VI. REFERENCES

1. Hariri AR, Tessitore A, Mattay VS, Fera F, Weinberger DR. The Amygdala Response to Emotional Stimuli: A Comparison of Faces and Scenes. *NeuroImage*. 2002;17(1):317-323.

doi:10.1006/nimg.2002.1179.
2. Kirschbaum C, Pirke KM, Hellhammer DH. The "Trier Social Stress Test" – A Tool for Investigating Psychobiological Stress Responses in a Laboratory Setting. *Neuropsychobiology*. 1993;28(1-2):76-81. doi:10.1159/000119004.
3. Allen AP, Kennedy PJ, Cryan JF, Dinan TG, Clarke G. Biological and psychological markers of stress in humans: Focus on the Trier Social Stress Test. *Neuroscience & Biobehavioral Reviews*. 2014;38:94-124. doi:10.1016/j.neubiorev.2013.11.005.
4. Edelson JT, Robertson GL. The effect of the cold pressor test on vasopressin secretion in man. *Psychoneuroendocrinology*. 1986;11(3):307-316. doi:10.1016/0306-4530(86)90016-8.
5. Velasco M, Gómez J, Blanco M, Rodriguez I. The cold pressor test: pharmacological and therapeutic aspects. *American journal of therapeutics*. 1997;4(1):34-38.
6. Rolke R, Baron R, Maier C, et al. Quantitative sensory testing in the German Research Network on Neuropathic Pain (DFNS): Standardized protocol and reference values. *Pain*. 2006;123(3):231-243. doi:10.1016/j.pain.2006.01.041.
7. Gorman JM, Papp LA, Coplan JD, et al. Anxiogenic effects of CO2 and hyperventilation in patients with panic disorder. *American Journal of Psychiatry*. April 1994. doi:10.1176/ajp.151.4.547.
8. Poma SZ, Milleri S, Squassante L, et al. Characterization of a 7% carbon dioxide (CO2) inhalation paradigm to evoke anxiety symptoms in healthy subjects. *Journal of Psychopharmacology*. 2005;19(5):494-503. doi:10.1177/0269881105056533.
9. Buss KA, Schumacher JRM, Dolski I, Kalin NH, Goldsmith HH, Davidson RJ. Right frontal brain activity, cortisol, and withdrawal behavior in 6-month-old infants. *Behavioral Neuroscience*. 2003;117(1):11. doi:10.1037/0735-7044.117.1.11.
10. Pfeifer M, Goldsmith HH, Davidson RJ, Rickman M. Continuity and Change in Inhibited and Uninhibited Children. *Child Development*. 2002;73(5):1474-1485. doi:10.1111/1467-8624.00484.
11. Norrholm SD, Vervliet B, Jovanovic T, et al. Timing of extinction relative to acquisition: A parametric analysis of fear extinction in humans. *Behavioral Neuroscience*. 2008;122(5):1016. doi:10.1037/a0012604.
12. Zeidan MA, Lebron Milad K, Thompson Hollands J, et al. Test–Retest Reliability during Fear Acquisition and Fear Extinction in Humans. *CNS Neuroscience & Therapeutics*. 2012;18(4):313-317. doi:10.1111/j.1755-5949.2011.00238.x.
13. Schmitz A, Grillon C. Assessing fear and anxiety in humans using the threat of predictable and unpredictable aversive events (the NPU-threat test). *Nature Protocols*. 2012;7(3):527-532. doi:10.1038/nprot.2012.001.
14. Samson AC, Kreibig SD, Soderstrom B, Wade AA, Gross JJ. Eliciting positive, negative and mixed emotional states: A film library for affective scientists. *Cognition & Emotion*. May 2015. doi:10.1080/02699931.2015.1031089.
15. Joormann J, Talbot L, Gotlib IH. Biased processing of emotional information in girls at risk for depression. *Journal of Abnormal Psychology*. 2007;116(1):135. doi:10.1037/0021-843X.116.1.135.
16. Cherek DR. Effects of smoking different doses of nicotine on human aggressive behavior. *Psychopharmacology*. 1981;75(4):339-345. doi:10.1007/BF00435849.

17. Geniole SN, MacDonell ET, McCormick CM. The Point Subtraction Aggression Paradigm as a laboratory tool for investigating the neuroendocrinology of aggression and competition. *Hormones and Behavior*. April 2016. doi:10.1016/j.yhbeh.2016.04.006.
18. Gagne JR, Van Hulle CA, Aksan N, Essex MJ, Goldsmith HH. Deriving childhood temperament measures from emotion-eliciting behavioral episodes: Scale construction and initial validation. *Psychological Assessment*. 2011;23(2):337. doi:10.1037/a0021746.
19. Wolpe J. *The Practice of Behavior Therapy*. Allyn & Bacon; 1990.
20. Kaplan DM, Smith T, Coons J. A validity study of the subjective unit of discomfort (SUD) score. *Measurement and Evaluation in Counseling and Development*. January 1995.
21. Wolpe J, Lang P. *Wolpe: Manual for the Fear Survey Schedule*. San Diego, CA: Educational and Industrial Testing Service; 1977.
22. Carleton RN, Norton MAPJ, Asmundson GJG. Fearing the unknown: A short version of the Intolerance of Uncertainty Scale. *Journal of Anxiety Disorders*. 2007;21(1):105-117. doi:10.1016/j.janxdis.2006.03.014.
23. Carver CS, White TL. Behavioral inhibition, behavioral activation, and affective responses to impending reward and punishment: The BIS/BAS Scales. *Journal of Personality and Social Psychology*. 1994;67(2):319. doi:10.1037/0022-3514.67.2.319.
24. Watson D, Friend R. Measurement of social-evaluative anxiety. 1969. doi:10.1111/j.2044-8341.1967.tb00577.x/abstract.
25. Taylor S, Zvolensky MJ, Cox BJ, et al. Robust dimensions of anxiety sensitivity: Development and initial validation of the Anxiety Sensitivity Index-3. *Psychological Assessment*. 2007;19(2):176. doi:10.1037/1040-3590.19.2.176.
26. Brown GW, Harris T. Social origins of depression: a reply. *Psychological Medicine*. 1978;8(04):577-588. doi:10.1017/S0033291700018791.
27. RUDOLPH KD, FLYNN M. Childhood adversity and youth depression: Influence of gender and pubertal status. *Development and Psychopathology*. 2007;19(02):497-521. doi:10.1017/S0954579407070241.
28. Bernstein DP, Fink L. *Bernstein: Childhood Trauma Questionnaire: a Retrospectiv... - Google Scholar*. 1998.
29. Ippen CG, Ford J, Racusin R, et al. *Traumatic Events Screening Inventory - Parent Report Revised*. 2002.
30. Taylor SE, Lerner JS, Sage RM, Lehman BJ, Seeman TE. Early Environment, Emotions, Responses to Stress, and Health. *Journal of Personality*. 2004;72(6):1365-1394. doi:10.1111/j.1467-6494.2004.00300.x.
31. Slavich GM, Epel ES. The Stress and Adversity Inventory (STRAIN): An automated system for assessing cumulative stress exposure. *Los Angeles: University of California*. 2010.
32. Wright KA, Lam DH, Brown RG. Reduced approach motivation following nonreward: Extension of the BIS/BAS scales. *Personality and Individual Differences*. 2009;47(7):753-757. doi:10.1016/j.paid.2009.06.015.
33. Baars MY, 252 M, ller MJ, et al. Depressive and Aggressive Responses to Frustration: Development of a Questionnaire and Its Validation in a Sample of Male Alcoholics. *Depression Research and Treatment*. 2011;2011. doi:10.1155/2011/352048.

Appendix NVS-I: NVS domain group table on task criteria for recommended paradigms

NVS Construct	Task	Valid Measure of Construct (#1)	Test-Retest Reliability (#2)	Floor/Ceiling (#2)	Practice Effects (#2)	Longitudinal Stability (#2)	Standardized Parameters (#4)	Copyright (#15)	Can it be used across many sites? (#9)	Can it be used with children or special populations? (#6)	Sensitivity to w/in person change (#10)	Tolerability (#18)
Acute threat	Trier Social Stress Test	4	*3*	*3*	?	?	2	5	5	Ages: 5 Culture: 4?	¾*	3
Acute threat	Behavioral approach Test	4	?	?	N/A	?	2	5	5	Ages:5 Culture: 4	5	4
Acute threat	Cold pressor and other pain tolerance tasks	4	3	?	?	3	4 (but not always attended to in literature)	5	3	Ages: 5 Culture: 5	3	3
Acute threat	CO^2 challenge	4	3	?			4	5	5	Ages: 5 Culture: 4	4	3
Acute threat	Stranger tasks	4	?	5	1	3	2	5	5	Ages: 5 Culture: 5	1	3
Acute threat (learning)	Fear conditioning	5	2	?	?	?	4	5	5	Ages: 5 Cultures:	5	4
Potential threat	NPU Threat-Task	4/5	4	5	?	?	2	5	3	Ages: 5 Culture: 3	?	3

Loss	Sadness eliciting film clips (see caveats above)	4	?	?	?	?	3	3 (some may be copyrighted)	5	Ages: 5 Culture: 4	3	4
Frustrative nonreward	LabTAB: Box Empty, Transparent Box	5	4	5	?	4 (only if mult. tasks are used)	2	5	3	5, Adults (potentially adaptable) Culture: ?	1	4
Frustrative nonreward	Points Subtraction Aggression Paradigm (PSAP)	4	1*	1*	?	1?	1	5	3	1 Could be adapted	?	4

NVS Construct (rest of criteria)	Task (rest of criteria)	Alternate forms (#2)	Internal reliability (#2)	Can be used in clinical trials or screening (#5)	Stand alone behavioral task? (#7)	Can use with imaging or ERP? (#11)	Relat btw task perf & neural sig known? (#13)	Clinical phenotype (#14)	Measures one construct or specific to one? (# 16)	Any task that could be modified? (#17)	Norms (#8)	Animal analogue (#12)
Acute threat	Trier Social Stress Test	1	??	1	5	4*	2	5	4	--	1	1
Threat	Behavioral approach	1	?	4	5	5	5	5	5	Lots of modifications	1	5
Acute threat	Cold Pressor and other pain tolerance tasks	1	?	2	5	5	5	4	1	Potentially	3	2

Acute threat	CO^2 challenge	1	N/A	4	5	5	4	5	5	Not really	1	4
Acute threat	Stranger tasks	1	4	2	5	2	2	2	3	?	1	5
Acute threat (learning)	Fear conditioning	5		4	5	5	4	5	4	Many	N/A	5
Potential threat	NPU		4	3 (with SUDS ratings)	3	5	5	5	5	--	1	5
Loss	Sadness eliciting film clips, esp w/immersion; facial exp or mood ratings	4	?	4	5	5	3	4	2/3	Yes, many	3	1
Frustrative nonreward	Points Subtraction Aggression Paradigm (PSAP)	1	?	4	5	5	3	4	3 (construct itself is broad. Measure taps loss as well as threat)	Could be modified for children	1	3
Frustrative nonreward	LabTAB: Box Empty, Transparent Box	1	4	3	5	2	--	3	3 (construct itself is broad. Measure taps cognitive control and positive emotionality)	Could be modified for adults	1	2

Appendix NVS-II: NVS domain group table on task criteria for paradigms that need more work, rated from 1 (no evidence) - 5 (strong evidence)

Construct	Task	Valid Measure of Construct (#1)	Test-Retest Reliability (#3)	Floor/Ceiling (#2)	Practice Effects (#2)	Longitudinal Stability (#2)	Standardized Parameters (#4)	Copyright (#15)	Can it be used across many sites? (#9)	Can it be used with kids or special	w/in person (#10)	Tolerability (#18)
Vigilance or Attentional capture	Dot Probe (supraliminal 1000ms presentation of angry/fearful faces)	3	2?	5	?	2?	5	5	5	5	5	5
Inability to disengage from negative stimulus	Exogenous Cuing Task	3 (psychometrics for non-emotional stimuli is available, but unclear if similar for emotional stimuli)		5	?		2	5	5	4	?	5

Construct (rest of criteria)	Task (rest of criteria)	Alternate forms (#2)	Internal reliability (#3)	Can be used in clinical trials	Stand alone behavioral tasks (#7)	Can use with imaging or	Relat btw task perf & neural sig known?	Clinical phenotype (#14)	Measures one construct or specific to	Any task that could be modified?	Norms (#8)	Animal analogue (#12)
Vigilance or Attentional capture	Dot Probe (supraliminal 1000ms presentation of angry/fearful faces)	2	1	4	5	5	4	5	3	--	1	3
Inability to disengage from negative stimulus	Exogenous Cuing Task	2	?	4	5	3	?	3	3	Lots of modif. are possible	1	1

Positive Valence Systems Final Report

Mauricio R. Delgado, Ph.D., Paul W. Glimcher, Ph.D. Greg Hajcak, Ph.D., Diego A. Pizzagalli, M.D., Ph.D. (chair), Michael T. Treadway, Ph.D., Benjamin E. Yerys, Ph.D.

Executive Summary

The Positive Valence Systems (PVS) Domain subgroup carefully considered the original PVS constructs and discussed over 25 tasks that were ranked according to pre-defined criteria. During deliberations concerning the PVS structure and possible reconfigurations, particular emphasis was placed on (1) avoiding potential redundancies across constructs/sub-constructs, and (2) attempting to isolate "purer" constructs. Similarly, when considering tasks, emphasis was placed on paradigms that isolate given sub-constructs. These considerations led to a proposed restructuring of the PVS domain into three constructs (*Reward Responsiveness*, *Reward Learning*, *Reward Valuation*), each involving three sub-constructs. Among all tasks discussed, 16 were selected for potential prioritization. Other tasks were discussed but not recommended, whereas others were deemed promising but requiring more evaluation.

I. General Comments

Before discussions of the current PVS structure and potential tasks, the workgroup deliberated on several general points:

1. Many tasks that have been adopted widely in the literature, particularly those developed for neuroimaging/neuropsychological studies in the late 1990s/early 2000s, often cannot disentangle current PVS constructs/sub-constructs. Prominent examples are the Iowa Gambling Task[1] and the Monetary Incentive Delay (MID) task [2], which are among the most widely used tasks in the field, and have provided a wealth of valuable information. For example, in the MID task, reward value and reward prediction error (RPE) are perfectly correlated, and thus cannot be de-conflated.
2. The MID task currently appears in the matrix for the construct Initial Responsiveness to Reward, however outcomes from this task that measure response to reward cannot be dissociated from each other. The PVS group is not recommending this task as a measure of Initial Responsiveness to Reward. Alternatively, they are recommending it as a measure of Reward Anticipation because the outcomes associated with anticipation are independent and can be isolated.
3. Although workgroup members deemed regulatory processes as being very important, there was consensus that regulation of PVS constructs would entail processes better captured by the Cognitive Systems and Arousal Systems. Accordingly, tasks probing regulatory processes were not discussed.

4. For some sub-constructs (in particular, "initial responsiveness to reward"), tasks yield no direct behavioral output. However, these constructs can be meaningfully probed with imaging, electrophysiological, and peripheral psychophysiological techniques, and could be augmented by affective ratings. Given the centrality of these constructs and their translational value for preclinical models, the absence of direct behavioral output did not prevent the recommendation of various tasks. In general, the ability of a task to be used in conjunction with imaging, electrophysiology or psychophysiology was deemed a plus.
5. Some constructs (e.g., Reward Prediction Error) require computational modeling for meaningful interpretation. Accordingly, dissemination of some of the proposed tasks within this sub-construct might be contingent upon (and thus limited by) expertise in computational modeling.
6. Although workgroup members acknowledged the importance of self-report measures of PVS constructs, performance-based or behavioral tasks were prioritized to maximize potential translation to and back-translation from preclinical (animal) models.
7. When evaluating tasks, tolerability (i.e., participants' experience) was also considered.

II. Organization of the Domain

During deliberations concerning the original PVS constructs and structure, particular emphasis was placed on (1) avoiding potential redundancies across constructs and sub-constructs, and (2) attempting to isolate "purer" constructs. These considerations led to a proposed restructuring of the PVS domain into three constructs (*Reward Responsiveness*, *Reward Learning*, *Reward Valuation*), each involving three sub-constructs (**Table 1**). Rationales for restructuring/renaming as well as de-prioritization of some task nominations from the original RDoC workshop are provided in later sections of this report.

TABLE 1: Proposed Restructuring of the PVS domain

Construct	Sub-construct
1. Reward Responsiveness	
	1.1. Initial Response to Reward
	1.2. Reward Anticipation
	1.3. Reward Satiation
2. Reward Learning	
	2.1. Habit

2.2. Probabilistic and Reinforcement Learning

2.3. Reward Prediction Error

3. Reward Valuation

3.1. Reward (probability)

3.2. Delay

3.3. Effort

III. Recommended Tasks

Paradigms that were evaluated as "best in class" for given PVS sub-constructs are described in more detail in **Appendices PVS-IIa-PVS-IIf**. Ratings on each suggested criterion for these tasks are provided in **Appendix PVS-I**.

1. Reward Responsiveness

We propose *Reward Responsiveness* as a construct, which includes three sub-constructs: initial responsiveness to reward, reward anticipation, and reward satiation.

1.1. Initial responsiveness to reward: This is defined by neural and physiological response to positive reinforcers (money, positive pictures). As such, by definition there are no optimal behavioral measures, at least not in current instantiations.

Guessing Task (e.g., Card Guessing or Doors) [3,4]. These tasks have no meaningful behavioral output; rather, they have been widely used in conjunction with e.g. fMRI, EEG/ERP, HR, GSR recordings. They could be modified to include ratings. These tasks have excellent construct validity and psychometric properties, although more work is needed to evaluate test-retest reliability. There are good data on individual differences and sensitivity to change. See **Appendix PVS-IIa** for detailed evaluations.

1.2. Reward anticipation

Monetary Incentive Delay Task[2]. Probes reward anticipation; modifications are needed in order to improve its ability to isolate anticipation; in particular use of longer and jittered interstimulus intervals (e.g., following an exponential function with over-representation of shorter inter-stimulus intervals) is expected to improve the ability to isolate anticipation-related activation. See **Appendix PVS-IIb** for detailed evaluations for each criterion for this task.

1.3. Sustained responsiveness to reward: We suggested renaming "Reward Satiation".

Fixed-Ratio Satiation Schedule [5]. Excellent construct validity, and potentially would have excellent other criteria, but needs significant development. This remains a recommendation due to the lack of a better option for this subconstruct.

2. Reward Learning

We propose *Reward Learning* as a construct, with sub-constructs of habit, reward prediction error, and probabilistic and reinforcement learning.

2.1. Habit

Devaluation Task [6,7]. This task has excellent construct validity. There are some concerns about ability to repeatedly administer (i.e., practice effects) and the use of the task with children/special populations. There is some evidence for links to clinical features. The psychometric properties are not yet known, but there are no better options for recommendation. See **Appendix PVS-IIc** for detailed evaluations for each criterion for this task.

Habit Task (longer term reversal learning; [8]. This task potentially has excellent construct validity but information about other parameters is unknown.

Habit Learning Task [9]. This task has excellent construct validity but information about other parameters is unkown. The task may not be practical or efficient because of length. Some proposed modifications include administering the task in only one session so that it could become more sensitive to individual differences.

2.2. Probabilistic and Reinforcement Learning (former Reward Learning)

Probabilistic Reward Task [10]. This task has excellent construct validity and acceptable test-retest, but there is a need to evaluate internal reliability (e.g., compute reliability for odd/even trials). It can be repeated, and used across many age and populations. It is sensitive to within-person change and has known relations to clinical features. Performance on this task can be manipulated by pharmacological (e.g., dopaminergic compounds) or behavioral (e.g., acute stressors) means in predictable manners. There are some emerging normative data for this task. See **Appendix PVS-IId** for detailed evaluations for each criterion for this task.

Pavlovian Conditioning[11]. In these tasks, one stimulus predicts a positive outcome. In spite of strong construct validity, these tasks yield poor behavioral profiles. Thus, they require imaging or psychophysiological (e.g., skin conductance, pupil dilation) readouts. Such tasks could be modified in order to include affective ratings. These tasks have unknown psychometric properties. They can be repeated, and used across many age and populations. There is evidence for some links to clinical features. There are no normative data for these tasks.

Probabilistic Stimulus Selection Task [12]. This task has excellent construct validity. Some concerns about this task include the fact that many participants do not learn it, it needs work from other labs using the task besides the initial lab, and it would be difficult to use with children or special populations. This task can be repeated. There is some evidence for links with clinical features. There are no normative data for this task.

Drifting Double Bandit [13] . This task has excellent construct validity, but unknown psychometrics properties. There is little evidence about links to clinical features. The sensitivity of the task to within-person change or clinical features is unknown. This task can be repeated if different stimulus sets are developed, and it can be used across many age and populations. There are no normative data for this task.

2.3. Expectancy/Reward Prediction Error

Rutledge Passive Lottery Task[14]. There are no behavioral outputs for this task, but it is a pure measure of RPE. The potential downside to this task is that modeling the data requires expertise in computational modeling.

Drifting Double Bandit [13]. This task has excellent construct validity, but unknown psychometrics properties. There is little evidence about links to clinical features. The sensitivity of these task to within person change or clinical features is unknown. This task can be repeated if different stimulus sets are developed, and it can be used across many age and populations. There are no normative data for this task.

3. Valuation

We argue for a construct called *Valuation*, with sub-constructs of Reward (which will encompass probability), delay, and effort.

3.1. Reward

Probability Choice Task [15] or analogous—drop ambiguity). See **Appendix PVS-IIe** for detailed evaluations.

Measuring the value subjects place on a reward in a way that allows inter-individual comparison based only on behavior is a theoretically difficult prospect. Using only behavior one can only measure difference in the "rate at which" the subjective value of a reward grows as a function of the rate at which the objective magnitude of a given reward grows. This is, formally, the curvature of the utility function from economics. Typically, the utility function is measured by asking questions that compete a fixed sized reward against rewards or greater magnitude but lower probability. Measurements of this type are very well developed in psychology and behavioral economics and typically are derivative of the classic Holt and Laurie [16] approach.

Willingness to Pay (BDM) [17] ; see also [18]

One way to begin to compare individual responses to rewards is to ask subjects to price in dollars (or in another currency) the maximum amount that they would be willing to pay to obtain a specific good under a specific condition. Two issues, however, make that measurement problematic. First, a subject who is actually bidding on a real good is often incentivized to report a low number in the hopes that they will 'game' the experimenter into giving them that good for less. This first concern is largely eliminated in the 'BDM' method. In the BDM method, subjects state the maximal price that they would be willing to pay from a menu of possible prices – say $1-$5 in 50 cent increments. Once they have reported that price they draw a chip from an urn with each chip bearing a single price from $1 - $5 in 50 cent increments. If the drawn chip is below their pre-stated maximum they buy the good for the price on the chip, if it is above the pre-stated maximum they are not allowed to buy the good. Under this regime the subjects do best if they report the true price because the pricing mechanism is unaffected by their 'bid'. BDM is for this reason the gold standard for assessing truly held 'values' (in dollars) for non-monetary goods. The second problem is that subjects should never be willing to pay more than the market price for a good if they can leave the lab immediately to purchase it for less. Typically, this is dealt with by asking the subjects to remain in the lab after bidding for some fixed length of time. For rare goods or goods with high market prices this is much less of a problem.

3.2. Delay

Workgroup deliberations as well as consideration of suggestions provided in response to the RFI highlight several candidate tasks, which have common features: Kable's task [19] (most often used in clinical samples), Traditional Bickel Hypothetical [20] (most often used in substance abuse literature), Johnson and Bickel[21], Green and Meyerson's hypothetical [22] (most often used in psychology studies). The task by Kable was deemed optimal for use with neuroimaging due to its display technique.

3.3. Effort valuation/willingness to work.

Effort Expenditure for Reward Task[23]. This task has good construct validity. It can be used in a range of populations, is sensitive to within-subject manipulations and can be used with children (> 9 years old, although without probability manipulation). The task has moderate to excellent test-retest reliability. Some minor concerns about this task were about about whether effort is confounded with time on task, but it was felt that a "pure" version could be developed by fixing trial timing structure (so time on task is held constant). See **Appendix PVS-IIe** for detailed evaluations for each criterion for this task.

TABLE 2: Recommended Tasks for each PVS sub-construct

Construct/Sub-construct	Task	Key references
1. Reward Responsiveness		
1.1. Initial Response to Reward	Simple Guessing Task	(Delgado et al. 2000)[2]
		(Carlson et al. 2011)[3]
1.2. Reward Anticipation	Monetary Incentive Delay Task	(Knutson et al. 2000)[1]
1.3. Reward Satiation	Fixed-ratio Satiation Schedule	(Sherman & Thomas 1968)[4]
2. Reward Learning		
2.1. Habit	Devaluation Task	(Gillan et al. 2011)[22]
	Habit Task	(McKim et al. 2016)[7]
	Habit Learning Task	(Tricomi et al. 2009)[8]
2.2. Probabilistic and Reinforcement Learning	Probabilistic Reward Task	(Pizzagalli et al. 2005)[9]
	Pavlovian Conditioning	(O'Doherty et al. 2004)[23]
	Drifting double bandit	(Daw et al. 2011)[11]
	Probabilistic Stimulus Selection Task	(Frank et al. 2004)[10]

2.3. Reward Prediction Error	Rutledge Passive Lottery Task	(Rutledge et al. 2010)[12]
	Drifting double bandit	(Daw et al. 2011)[11]
3. Reward Valuation		
3.1. Reward (probability)	Probability Choice Task	[e](Levy et al. 2010)[13]
	Willingness To Pay Task	(Becker et al. 1963)[15]
3.2. Delay	Delayed Discounting Task	[f](Kable & Glimcher 2007)[17]
		(Johnson & Bickel 2002)[19]
		(Green & Myerson 2004)[20]
3.3. Effort	Effort Expenditure for Reward Task	(Treadway et al. 2009)[21]

[e] Drop ambiguity manipulation

[f] Deemed preferable in conjunction with functional neuroimaging

IV. Tasks that require more evaluation

Reward Valuation

3.3. Effort

a) Deck Choice Effort Task[24] : This cognitive effort-based decision making task was developed for use in clinical populations[24]. The Deck task involves making choices between hard vs. easy cognitive tasks (i.e., cognitive set switching) for different levels of monetary reward. It is based on a cognitive effort task originally developed for healthy individuals[25]. The construct of cognitive effort has been studied in animal models[26]; this task was nominated in response to NIMH RFI posted on Monday 3/28/2016 (https://grants.nih.gov/grants/guide/notice-files/NOT-MH-16-007.html).

- *Evidence for construct validity in terms of the mechanism the test is thought to assess:* Evidence for construct validity stems from: its ability to distinguish schizophrenia patients from healthy subjects, relations to other effort-based decision making tasks in schizophrenia, non-clinical research showing some convergent validity for different types of cognitive effort tasks, and neuroimaging studies.
- *Evidence for reliability, of any form, including internal consistency, test-retest reliability, etc.:* Modest-to-Good (ICC = .67) one-month test-retest reliability in patients with schizophreniar[24].
- *Evidence for other relevant psychometric characteristics about the test, including practice effects, floor or ceiling effects, etc.*: The task performed reasonably well regarding floor, ceiling, or practice effects in schizophrenia[24].
- *Descriptions of any known animal homologues for this test*: Cognitive effort-based decision making tasks have been used in animal models[26].
- *Evidence of task improvement with psychological or pharmacological treatment:* None

b) Cognitive Effort Discounting (COGED) task[27]: The COGED is used to assess evaluation of cognitive effort costs, balanced against rewards. The extent to which an individual discounts a reward, contingent on performance of a demanding task, is thought to indicate how strongly they experience effort costs in the cognitive domain, and conversely, their motivation for goal pursuit via cognitive engagement.

- *Evidence for construct validity in terms of the mechanism the test is thought to assess:* COGED is sensitive to both state and trait factors that support its construct validity. State factors include working memory load ('N' on the N-back task), which increases discounting and offer amount, which decreases discounting[27]. Trait factors include Need for Cognition and cognitive aging[27] and negative symptoms in schizophrenia[28]. Moreover, unpublished observations indicate that COGED is strongly correlated with switch costs in a tasks-switching paradigm (steep discounters on the N-back have larger switch costs in a different task-switching paradigm), and a weaker correlation with delay discounting (steep effort discounters are almost invariably steep/impatient delay discounters). This latter observation dovetails with recent studies of cognitive effects in delay discounting

supporting that patient choice behavior requires (potentially effortful) working memory allocation during decision-making.

Finally, at a neural level, recent observations (in preparation for publication) include that dimensions of reward amount and task load are both robustly encoded in canonical subjective value encoding regions like the vmPFC and posterior cingulate cortex, as participants evaluate cognitive effort-contingent rewards. Also, while participants are engaged with the N-back task, steeper effort discounters show greater recruitment in a number of task-positive regions including the fronto-parietal, salience, and dorsal attention networks.

- *Evidence for reliability, of any form, including internal consistency, test-retest reliability, etc.:* Limited evidence, but includes the aforementioned inter-individual correlations between COGED and Need for Cognition and negative symptoms. Correlations with these trait dimensions support reliability. In a small sample (N = 25 participants), the ICC of the Area Under the Discounting Curve measure of COGED, across three sessions among healthy young adults, was 0.47 with 95% CI of [0.23, 0.69]. To the extent that COGED captures both trait and state effects (e.g. fatigue or sleep deprivation[29], some variability is expected.
- *Evidence for other relevant psychometric characteristics about the test, including practice effects, floor or ceiling effects, etc.*: Depending on paradigm design, brief exposure to the N-back task prior to discounting yields shallower discounting than prolonged exposure which, at the limit, produces no discounting (individuals always select the more demanding option for more money), restricting inter-individual variability.
- *Descriptions of any known animal homologues for this test*: The nearest is the Rat Cognitive Effort Task (RCET) of Cocker and colleagues[30]. There are many other physical effort paradigms (e.g. T-mazes for rats, or level pulls for monkeys), but this is the only animal *cognitive* effort task.
- *Evidence of task improvement with psychological or pharmacological treatment:* None is available to date.

c) Additional Effort-based Tasks

The following tasks were deemed as promising in light of their potential ability to probe particular sub-constructs but require more work and evaluation (often because they have been investigated in a limited number of studies):

- Physical Effort: Grip Force Task [31,32]
- Physical Effort: Beautiful Faces Task [33] and related tasks (e.g., to probe restricted interest in autism[34]): Tasks that require effort to experience a reinforcer (e.g., beautiful faces) need to be refined to better index effort.

V. Tasks that are not recommended

1. Reward Responsiveness

1.1. Initial Response to Reward

Monetary Incentive Delay Task[2]: The MID was recommended for the sub-construct "Reward Anticipation" (see Section III, Point 1.2.) but not for the sub-construct "Initial Responsiveness to Reward" due to poor validity in dissociating prediction error and outcome value signals. Specifically, formal modeling expectation of reward outcome in this task is challenging, as it will be influenced by both outcomes of prior trials as well as performance on the current trial (e.g., depending on RT to the target, participants have a good expectation of the upcoming outcome). Moreover, common attempts to work around this limitation by modeling expectation as just the average reward rate result in prediction error values that are co-linear with outcome values.

Cue Reactivity Tasks: Not considered because these tasks likely engage reward anticipation in some conditions, but initial responsiveness to reward in others.

1.2. Reward Anticipation

None.

1.3. Reward Satiation

Devaluation Tasks: Poor construct validity for this sub-construct—likely a much better measure of habit.

2. Reward Learning

2.1. Habit

Knot Tying and Serial Response Tasks: not discussed because they mostly probe procedural learning; similarly, Attentional Blindness Tasks were not considered because they assess attentional bias related to expertise.

2.2. Probabilistic and Reinforcement Learning

Drifting Bandit [14] *:* Excellent properties; it has been used in studies in Parkinson and dopaminergic challenges; it can measure exploration/exploitation and allows to fit learning rate and bias. It was not discussed further, however, because Double Bandit Tasks (e.g., Daw et al. 2011) yield the same outcome variables and allow one to parse both model-based and model-free parameters in the same task, and are thus more efficient.

Pavlovian Instrumental Transfer Tasks[35]: interesting task but it requires work before wide dissemination because ~30% of participants are unable to learn it).

2.3. Reward Prediction Error

None.

3. Reward Valuation

3.1. Reward (probability)

None.

3.2. Delay

None.

3.3. Effort

Progressive Ratio Task[36]*:* In spite of their widespread use in the literature, Progressive Ratio Tasks were not considered because they confound effort, time discounting, reward magnitude (and satiety).

VI. References

1. Bechara A, Damasio H, Tranel D, Damasio AR. Deciding Advantageously Before Knowing the Advantageous Strategy. *Science*. 1997;275(5304):1293-1295. doi:10.1126/science.275.5304.1293.
2. Knutson B, Westdorp A, Kaiser E, Hommer D. FMRI Visualization of Brain Activity during a Monetary Incentive Delay Task. *NeuroImage*. 2000;12(1):20-27. doi:10.1006/nimg.2000.0593.
3. Delgado MR, Nystrom LE, Fissell C, Noll DC, Fiez JA. Tracking the Hemodynamic Responses to Reward and Punishment in the Striatum. *Journal of Neurophysiology*. 2000;84(6):3072-3077. doi:10.1016/0166-2236(90)90107-L.
4. Carlson JM, Foti D, Mujica-Parodi LR, Harmon-Jones E, Hajcak G. Ventral striatal and medial prefrontal BOLD activation is correlated with reward-related electrocortical activity: A combined ERP and fMRI study. *NeuroImage*. 2011;57(4):1608-1616. doi:10.1016/j.neuroimage.2011.05.037.
5. Sherman JA, Thomas JR. Some Factors Controlling Preference Between Fixed-Ratio and Variable-Ratio Schedules of Reinforcement. *Journal of the Experimental Analysis of Behavior*. 1968;11(6):689–&.
6. Gottfried JA, O'Doherty J, Dolan RJ. Encoding Predictive Reward Value in Human Amygdala and Orbitofrontal Cortex. *Science*. 2003;301(5636):1104-1107. doi:10.1126/science.1087919.
7. Balleine B, Dickinson A. Signalling and incentive processes in instrumental reinforcer devaluation. *The Quarterly Journal of Experimental Psychology*. May 2007. doi:10.1080/14640749208401007.
8. McKim TH, Bauer DJ, Boettiger CA. Addiction history associates with the propensity to form habits. *Journal of Cognitive Neuroscience*. 2016;28(7):1024-1038.

doi:10.1162/jocn_a_00953.
9. Tricomi E, Balleine BW, O'Doherty JP. A specific role for posterior dorsolateral striatum in human habit learning. *European Journal of Neuroscience*. 2009;29(11):2225-2232. doi:10.1111/j.1460-9568.2009.06796.x.
10. Pizzagalli DA, Jahn AL, O'Shea JP. Toward an objective characterization of an anhedonic phenotype: A signal-detection approach. *Biological Psychiatry*. 2005;57(4):319-327. doi:10.1016/j.biopsych.2004.11.026.
11. O'Doherty J, Dayan P, Schultz J, Deichmann R, Friston K, Dolan RJ. Dissociable Roles of Ventral and Dorsal Striatum in Instrumental Conditioning. *Science*. 2004;304(5669):452-454. doi:10.1126/science.1094285.
12. Frank MJ, Seeberger LC, O'Reilly RC. By Carrot or by Stick: Cognitive Reinforcement Learning in Parkinsonism. *Science*. 2004;306(5703):1940-1943. doi:10.1126/science.1102941.
13. Daw ND, Gershman SJ, Seymour B, Dayan P, Dolan RJ. Model-Based Influences on Humans' Choices and Striatal Prediction Errors. *Neuron*. 2011;69(6):1204-1215. doi:10.1016/j.neuron.2011.02.027.
14. Rutledge RB, Dean M, Caplin A, Glimcher PW. Testing the reward prediction error hypothesis with an axiomatic model. *Journal of Neuroscience*. 2010;30(40):13525-13536. doi:10.1523/JNEUROSCI.1747-10.2010.
15. Levy I, Snell J, Nelson AJ, Rustichini A, Glimcher PW. Neural Representation of Subjective Value Under Risk and Ambiguity. *Journal of Neurophysiology*. 2010;103(2):1036-1047. doi:10.1152/jn.00853.2009.
16. Holt CA, Laury SK. Risk aversion and incentive effects. *American economic review*. 2002.
17. Becker GM, Degroot MH, Marschak J. Stochastic models of choice behavior. *Behavioral Science*. 1963;8(1):41-55. doi:10.1002/bs.3830080106.
18. Louie K, Khaw MW, Glimcher PW. Normalization is a general neural mechanism for context-dependent decision making. *PNAS*. 2013;110(15):6139-6144. doi:10.1073/pnas.1217854110.
19. Kable JW, Glimcher PW. The neural correlates of subjective value during intertemporal choice. *Nature neuroscience*. 2007;10(12):1625-1633. doi:10.1038/nn2007.
20. Bickel WK, Marsch LA. Toward a behavioral economic understanding of drug dependence: delay discounting processes. *Addiction*. 2001;96(1):73-86. doi:10.1046/j.1360-0443.2001.961736.x.
21. Johnson MW, Bickel WK. WITHIN-SUBJECT COMPARISON OF REAL AND HYPOTHETICAL MONEY REWARDS IN DELAY DISCOUNTING. *Journal of the Experimental Analysis of Behavior*. 2002;77(2):129-146. doi:10.1901/jeab.2002.77-129.
22. Green L, Myerson J. A Discounting Framework for Choice With Delayed and Probabilistic Rewards. *Psychological Bulletin*. 2004;130(5):769. doi:10.1037/0033-2909.130.5.769.
23. Treadway MT, Buckholtz JW, Schwartzman AN, Lambert WE, Zald DH. Worth the "EEfRT?" The effort expenditure for rewards task as an objective measure of motivation and anhedonia. *PLOS ONE*. 2009;4(8):e6598. doi:10.1371/journal.pone.0006598.
24. Reddy LF, Horan WP, Barch DM, et al. Effort-Based Decision-Making Paradigms for Clinical Trials in Schizophrenia: Part 1—Psychometric Characteristics of 5 Paradigms. *Schizophr Bull*. 2015;41(5):sbv089-sbv1054. doi:10.1093/schbul/sbv089.

25. Kool W, McGuire JT, Rosen ZB, Botvinick MM. Decision making and the avoidance of cognitive demand. *Journal of Experimental Psychology: General*. 2010;139(4):665. doi:10.1037/a0020198.
26. Young JW, Markou A. Translational Rodent Paradigms to Investigate Neuromechanisms Underlying Behaviors Relevant to Amotivation and Altered Reward Processing in Schizophrenia. *Schizophr Bull*. 2015;41(5):1024-1034. doi:10.1093/schbul/sbv093.
27. Westbrook A, Kester D, Braver TS. What Is the Subjective Cost of Cognitive Effort? Load, Trait, and Aging Effects Revealed by Economic Preference. *PLOS ONE*. 2013;8(7):e68210. doi:10.1371/journal.pone.0068210.
28. Culbreth A, Westbrook A, Barch D. Negative symptoms are associated with an increased subjective cost of cognitive effort. *Journal of Abnormal Psychology*. 2016;125(4):528. doi:10.1037/abn0000153.
29. Libedinsky C, Massar S, Ling A, Chee W, Huettel SA. Sleep deprivation alters effort discounting but not delay discounting of monetary rewards. *Sleep*. 2013.
30. Cocker PJ, Hosking JG, Benoit J, Winstanley CA. Sensitivity to Cognitive Effort Mediates Psychostimulant Effects on a Novel Rodent Cost|[sol]|Benefit Decision-Making Task. *Neuropsychopharmacology*. 2012;37(8):1825-1837. doi:10.1038/npp.2012.30.
31. Cléry-Melin M-L, Schmidt L, Lafargue G, Baup N, Fossati P, Pessiglione M. Why Don't You Try Harder? An Investigation of Effort Production in Major Depression. *PLOS ONE*. 2011;6(8):e23178. doi:10.1371/journal.pone.0023178.
32. Schmidt L, Lebreton M, Cléry-Melin M-L, Daunizeau J, Pessiglione M. Neural mechanisms underlying motivation of mental versus physical effort. *PLoS Biology*. 2012;10(2):e1001266. doi:10.1371/journal.pbio.1001266.
33. Aharon I, Etcoff N, Ariely D, Chabris CF, O'Connor E, Breiter HC. Beautiful Faces Have Variable Reward Value. *Neuron*. 2001;32(3):537-551. doi:10.1016/S0896-6273(01)00491-3.
34. Cascio CJ, Foss Feig JH, Heacock J, et al. Affective neural response to restricted interests in autism spectrum disorders. *Journal of Child Psychology and Psychiatry*. 2014;55(2):162-171. doi:10.1111/jcpp.12147.
35. Corbit LH, Balleine BW. Double Dissociation of Basolateral and Central Amygdala Lesions on the General and Outcome-Specific Forms of Pavlovian-Instrumental Transfer. *J Neurosci*. 2005;25(4):962-970. doi:10.1523/JNEUROSCI.4507-04.2005.
36. Hodos W, Kalman G. EFFECTS OF INCREMENT SIZE AND REINFORCER VOLUME ON PROGRESSIVE RATIO PERFORMANCE. *Journal of the Experimental Analysis of Behavior*. 1963;6(3):387-392. doi:10.1901/jeab.1963.6-387.

Appendix 1: Ratings of tasks recommended for consideration.

DOMAIN: POSITIVE VALENCE SYSTEMS

Construct	Task	Criteria (Rate each on a scale of 1-5, with 1 = does not do a good job of meeting the criterion; 5 = does an excellent job meeting the criterion Valid Measure of Construct	Test-Retest Reliability	Floor/Ceiling	Practice Effects	Longitudinal Stability	Sensitive to w/in person change	Standardized Parameters	Can it be used across sites?	Can it be used with kids or special populations?	Are normative data available?	Are relations to clinical features known?	Not copyrighted?
1. Reward Responsiveness													
1.1. Initial Response to Reward	Simple Guessing Task (e.g. 50% Card Task)	5[7]	3	5	5	4	5	3	5	4[8]	1	4	5

[7] But no behavioral outcome available. It would require the addition of self-report or psychophysiological assessments.

[8] With children than 7 years old

1.2. Reward Anticipation	Monetary Incentive Delay Task	1[9],2[10]	2[11]	5	4	1 (u)[12]	5	3	5	5	5	4	5
1.3 Reward Satiation	Fixed-ratio Satiation Schedule	5	1	4	5	1	1	1	4	4	1	1	5
2. Reward Learning													
2.1. Habit	Devaluation Task	5	1 (u)[f]	4	3	1 (u)[f]	1 (u)[f]	1	4	3[13]	1	1	5
	Habit Task												
	Habit Learning Task												
2.2. Probabilistic and Reinforcement Learning	a) Probabilistic Reward Task	5	3	5	5	4	5	4	5	4[b]	4	4	4[14]
	b) Pavlovian Conditioning	5[a]	1 (u)[f]	4	3	1 (u)[f]	3	3	5	5	1	3	5
	c) Drifting double bandit	5	1 (u)[f]	5	5	1 (u)[f]	2	4	5	4	1	1	5
	d) Probabilistic Stimulus Selection Task	5	1 (u)[f]	3	5	1 (u)[f]	4	4	3	3	1	4	5
2.3. Reward Prediction Error	a) Rutledge Passive Lottery Task	5	1 (u)[f] 1 (u)[f]	3	4	1 (u)[f] 1 (u)[f]	1 (u)[f]	3	2	3	1	1	5
	b) Drifting Double Bandit (see above)	5		5	5		2	4	5	4	1	1	5

[9] For behavioral outcome

[10] For neural outcome

[11] Good but low N

[12] Unknown

[13] Children, OCD, autism

[14] Freely available for researchers/non-profits; use by industry requires licensing

3. Reward Valuation													
3.1. Reward (probability)	Probability Choice Task	5	1 (u)[f] 1 (u)[f]	3	4	1 (u)[f] 1 (u)[f]	1 (u)[f]	3	2	3	1	1	5
	Willingness To Pay Task	5		3	3		1 (u)[f]	3	3	3	1	1	5
3.2. Delay	Delayed Discounting Task												
3.3. Effort	Effort Expenditure for Reward Task	4[15]	3	5	4	4	5	3	5	4	2	4	4[h]

[15] It manipulates efforts and timeline at the same time, thus not a pure measure of effort. Modifications could be applied.

Appendix PVS-IIa: Detailed Evaluation Criteria for Simple Guessing Task

PVS Construct: *Reward Responsiveness*

PVS Sub-construct: *Initial Response to Reward*

A) Card-Guessing Task (e.g., Delgado et al.)

1. How valid a test of the construct is the task?

The card guessing task was developed to identify neural circuits involved in reward processing. More specifically, the task allows for the comparison of brain responses to positive outcomes (e.g., receiving a monetary reward) compared to neutral or negative outcomes (e.g., monetary loss). The original version of the task [1] does not require learning or much practice and merely involves "guessing" decisions whether the number of a card is higher or lower than 5 (at a 50% probability), with guesses resulting in positive (a correct response), neutral or negative (an incorrect response) outcomes, thus controlling for changes in responses to reward as a function of learning or expectations. Prior studies have shown that:

(1) Across several paradigms, a comparison of positive and negative outcomes yields activation in reward-related regions, primarily dorsal and ventral striatum (for review see [2]). This can be observed in both event-related and blocked designs.
(2) This reward-related response is context-dependent and can be modulated by factors such as magnitude [3], probability [4] and the type of reward utilized, from non-monetary positive feedback [5, 6] to symbolic stimuli representing food [7].
(3) This reward-related response is blunted by exposure to acute stress [8] or deprivation of nicotine [9].
(4) This reward-related response characterized by the card-guessing task has been found to be altered in a population of patients recovered from anorexia nervosa [10] and bulimia nervosa [11] as well as adolescents with Anorexia Nervosa [12].
(5) This reward-related response characterized by the card-guessing task has also been found to be altered in a population of adolescents with major depressive disorder [13, 14], with such alterations being predictive of depressive symptoms in pubertal adolescents [15] or related to challenging social experiences in early adolescence (such as peer victimization; [16]).
(6) Reward-related responses characterized by the card-guessing task are susceptible to the social context in which they are received, being altered based on the perception

of a cooperative or competitive scenario [17, 18] or as a function of whether the interaction is with a person or computer [19, 20].

(7) Childhood measures of stress [21] and emotional neglect [22] correlate with blunted-reward sensitivity as measured by neural responses to rewarding outcomes in the card-guessing task.

(8) The magnitude of this reward-related response correlates with preferences for immediate over delayed rewards [23] and risky choices in some contexts [19], as well as an unwillingness to resist cigarette smoking [24].

(9) Sustained activity in reward-related regions during this paradigm in the laboratory correlates with real world positive emotional responses in control participants [25] and positive affect in adolescent major depressive disorder [14].

(10) The card-guessing task can yield results in long [1] or short (localizer; [26]) versions and can be modified to also look at anticipation of reward or changes as a function of learning or other factors (e.g., social context [18]).

2. **Does the task have good psychometric characteristics (incl. high internal reliability, test-retest reliability, sensitivity/specificity, limited practice effects, availability of alternate forms, longitudinal stability)?**
   i. high internal reliability: Not evaluated
   ii. test-retest reliability: Not evaluated
   iii. Sensitivity/specificity: Not evaluated
   iv. Limited practice effects: In original version, there are no known practice effects.
   v. Availability of alternate forms: Yes, there is a high degree of flexibility with this paradigm and it has been adapted for different questions, or timing constraints or for specific populations.
   vi. Longitudinal stability: Not evaluated

3. **Are parameters for administering the task (e.g., number of trials, stimulus characteristics, etc.) standardized on an empirical basis?**
   i. There are no formal empirical parameters as behavioral measures beyond reaction time and subjective measures that serve as manipulation checks are not included or optimal for analysis. There are published minimum amount of trials in various adaptations that have been effective.

4. **Is the task free from floor/ceiling effects which would preclude use in subjects with a range of impairment?**
   i. Yes

5. **Does the task have the same performance characteristics across cultures? Is it free from culture- and language-specific features/stimuli?**
   i. The task should be free from culture and language-specific features beyond changing the currency.
   ii. One example: the task has been run in Germany with adult ADHD participants to similar results (Wilbertz et al., 2012).

6. **Is the task sensitive to change and lack and loss of function?**
   i. Yes, as evidenced from its use in diverse patient populations previously described.

7. **Can the task (or its analog) be used in animals?**
   i. Yes, although there are no clear parallel tasks at this time.

8. **Can the task be used across age groups?**
   i. Yes, the task has been used with children, adolescents and older adults.

9. **Can the task be used with methods to interrogate brain circuitry (e.g., fMRI, EEG, etc.)?**
   i. The primary goal of this task is to observe reward-related activation. As such, it is used primarily with neuroimaging methods such as fMRI.

10. **Is there consensus on which metric/score should be considered to be primary?**
    i. The task serves primarily as a measure of neural activity of reward responses. Thus, the primary measure is a measure of BOLD signals in reward-related regions.

11. **Are adequate normative data available across age, gender, education, ethnicity, SES?**
    i. The data are available across multiple paradigms but have not been aggregated.

12. **Are the relationships between task performance and neural signal(s) known?**
    i. Yes (se point #1)

13. **Are the relationships between task performance and clinical feature(s) known?**
    i. Yes for some clinical features (see point #1)

14. **Is the task feasible for administration across sites?**
    i. Yes

15. **Can the task be used as a stand-alone behavioral task?**
    i. No

16. What work is needed to get this task ready for use in clinical trials?

i. Selection of one version of the paradigm that can be standardized (e.g., based on amount of trials and optimal timing).

17. Is the task copyrighted?

i. No

References

1. Delgado, M.R., et al., *Tracking the hemodynamic responses to reward and punishment in the striatum.* J Neurophysiol, 2000. **84**(6): p. 3072-7.
2. Delgado, M.R., *Reward-related responses in the human striatum.* Ann N Y Acad Sci, 2007. **1104**: p. 70-88.
3. Delgado, M.R., et al., *Dorsal striatum responses to reward and punishment: effects of valence and magnitude manipulations.* Cogn Affect Behav Neurosci, 2003. **3**(1): p. 27-38.
4. Delgado, M.R., et al., *An fMRI study of reward-related probability learning.* Neuroimage, 2005. **24**(3): p. 862-73.
5. Delgado, M.R., V.A. Stenger, and J.A. Fiez, *Motivation-dependent responses in the human caudate nucleus.* Cereb Cortex, 2004. **14**(9): p. 1022-30.
6. Tricomi, E., et al., *Performance feedback drives caudate activation in a phonological learning task.* J Cogn Neurosci, 2006. **18**(6): p. 1029-43.
7. Luking, K.R. and D.M. Barch, *Candy and the brain: neural response to candy gains and losses.* Cogn Affect Behav Neurosci, 2013. **13**(3): p. 437-51.
8. Porcelli, A.J., A.H. Lewis, and M.R. Delgado, *Acute stress influences neural circuits of reward processing.* Front Neurosci, 2012. **6**: p. 157.
9. Wilson, S.J., et al., *Effect of smoking opportunity on responses to monetary gain and loss in the caudate nucleus.* J Abnorm Psychol, 2008. **117**(2): p. 428-34.
10. Wagner, A., et al., *Altered reward processing in women recovered from anorexia nervosa.* Am J Psychiatry, 2007. **164**(12): p. 1842-9.
11. Wagner, A., et al., *Altered striatal response to reward in bulimia nervosa after recovery.* Int J Eat Disord, 2010. **43**(4): p. 289-94.
12. Bischoff-Grethe, A., et al., *Altered brain response to reward and punishment in adolescents with Anorexia nervosa.* Psychiatry Res, 2013. **214**(3): p. 331-40.
13. Forbes, E.E., *fMRI studies of reward processing in adolescent depression.* Neuropsychopharmacology, 2011. **36**(1): p. 372-3.
14. Forbes, E.E., et al., *Altered striatal activation predicting real-world positive affect in adolescent major depressive disorder.* Am J Psychiatry, 2009. **166**(1): p. 64-73.
15. Morgan, J.K., et al., *Neural response to reward as a predictor of increases in depressive symptoms in adolescence.* Neurobiol Dis, 2013. **52**: p. 66-74.

16. Casement, M.D., et al., *Girls' challenging social experiences in early adolescence predict neural response to rewards and depressive symptoms.* Dev Cogn Neurosci, 2014. **8**: p. 18-27.
17. Fareri, D.S. and M.R. Delgado, *Differential reward responses during competition against in- and out-of-network others.* Soc Cogn Affect Neurosci, 2014. **9**(4): p. 412-20.
18. Fareri, D.S., et al., *Social network modulation of reward-related signals.* J Neurosci, 2012. **32**(26): p. 9045-52.
19. Delgado, M.R., et al., *Understanding overbidding: using the neural circuitry of reward to design economic auctions.* Science, 2008. **321**(5897): p. 1849-52.
20. Fareri, D.S., L.J. Chang, and M.R. Delgado, *Computational substrates of social value in interpersonal collaboration.* J Neurosci, 2015. **35**(21): p. 8170-80.
21. Hanson, J.L., et al., *Cumulative stress in childhood is associated with blunted reward-related brain activity in adulthood.* Soc Cogn Affect Neurosci, 2016. **11**(3): p. 405-12.
22. Hanson, J.L., A.R. Hariri, and D.E. Williamson, *Blunted Ventral Striatum Development in Adolescence Reflects Emotional Neglect and Predicts Depressive Symptoms.* Biol Psychiatry, 2015. **78**(9): p. 598-605.
23. Hariri, A.R., et al., *Preference for immediate over delayed rewards is associated with magnitude of ventral striatal activity.* J Neurosci, 2006. **26**(51): p. 13213-7.
24. Wilson, S.J., et al., *Weak ventral striatal responses to monetary outcomes predict an unwillingness to resist cigarette smoking.* Cogn Affect Behav Neurosci, 2014. **14**(4): p. 1196-207.
25. Heller, A.S., et al., *The Neurodynamics of Affect in the Laboratory Predicts Persistence of Real-World Emotional Responses.* J Neurosci, 2015. **35**(29): p. 10503-9.
26. Speer, M.E., J.P. Bhanji, and M.R. Delgado, *Savoring the past: positive memories evoke value representations in the striatum.* Neuron, 2014. **84**(4): p. 847-56.

B. Doors Task (e.g., Hajcak et al.)

1. How valid a test of the construct is the task?

The Doors Task is simple gambling task that is intended to elicit physiological responses to receiving reward and loss. On each trial, participants view two doors and are told that one door leads to monetary reward and one lead to monetary loss; participants select a door by clicking the left or right mouse button, and subsequently receive feedback indicating either a win ($.50) or a loss ($.25). The task includes 60 trials, and feedback is exactly equiprobable (i.e., 30 gains and 30 losses, presented in a random order).

This task, and functionally identical variants like the balloon task or cards guessing task, was designed to examine physiological responses (i.e., EEG, fMRI) to favorable (i.e., winning money) versus unfavorable (i.e., losing money) feedback. The studies below focus on ERP response to reward and striatal response to reward—though other regions of interest have been examined in relation to individual differences. Prior studies have shown that neural response to reward is:

(1) Related across both ERP and fMRI methods[1,5,11].
(2) Related to behavioral measures of reward sensitivity and self-reported sensitivity to reward[2]. and real-world positive affective experience [8].
(3) Blunted in relation to increased depressive symptoms in both children and adults[2,4,9].

(4) Blunted among individuals with MDD, especially in relation to anhedonic symptoms[7,9,14]. One recent study found reduced reward response among remitted melancholic MDD individuals[20].
(5) Reduced among individuals at high risk for depression[12,21], and reduced reward-related brain activity predicts increases in depressive symptoms [2,4,16] and new-onset depression prospectively [2,16].
(6) Abnormal among individuals with addiction [17], especially in relation to anhedonic symptoms and predicted rewards.
(7) Is linked to genes that regulate DA[10].
(7) Correlated among first-degree relatives (r=0.31)[21].
(8) That is blunted in depression may improve with therapy [6].

2. Does the task have good psychometric characteristics (incl. high internal reliability, test-retest reliability, sensitivity/specificity, limited practice effects, availability of alternate forms, longitudinal stability)?

i. high internal reliability: high internal reliability for both striatum response to reward (r=0.66)[15] and ERP response to reward (r=0.85[15]; r=0.90[3]; r=0.89[13]).
ii. test-retest reliability: using fMRI – moderate (ICCs=0.55–0.62)[18]; using ERP, moderate-to-high (r=0.67[3]; r=0.71[13])
iii. Sensitivity/specificity: ROC analyses not performed yet (but data are available)
iv. Limited practice effects: task can be done many times.
v. Availability of alternate forms: Yes, doors task is functionally identical to card guessing and similar tasks where probability of reward is 50% on each trial[7,8].

3. Are parameters for administering the task (e.g., number of trials, stimulus characteristics, etc.) standardized on an empirical basis?

i. Task stimuli were optimized to be simple and can be used with a large age range (as low as 4 years in ongoing work); although the task produces internally reliable reward-related neural measures with 40-60 trials, it appears that half as many trials may be required[15]—though whether task length impacts relationships with individual differences is unknown.

4. Is the task free from floor/ceiling effects which would preclude use in subjects with a range of impairment?

i. Yes; task has been used with children and other special populations.

5. Does the task have the same performance characteristics across cultures? Is it free from culture- and language-specific features/stimuli?

i. In addition to the U.S., the task has been used in Asian samples, where reward-related neural activity has also been related to depression and anhedonia [14].
ii. There is no a priori reason to believe that it would perform in a culturally-specific way.

6. Is the task sensitive to change and lack and loss of function?

i. Yes, as evidenced by findings in patient samples (see above).

7. Can the task (or its analog) be used in animals?

i. In theory, yes; though an animal version has not been created.

8. Can the task be used across age groups?

i. Yes, published data in 9 year-olds[12]; ongoing work in 3-6 year olds.

9. Can the task be used with methods to interrogate brain circuitry (e.g., fMRI, EEG, etc.)?

i. Yes; the task has been used to examine reward-related brain activity using both EEG and fMRI.

10. Is there consensus on which metric/score should be considered to be primary?

i. Reward Positivity (in ERP), or the difference between reward and non-reward (i.e., average activity from 250-350 ms following feedback at FCz; this appears later among younger subjects)
ii. Reward-circuit activation using fMRI (i.e., striatal response, medial prefrontal cortex response)

11. Are adequate normative data available across age, gender, education, ethnicity, SES?

i. The task has been administered to more than 1,000 individuals. Age- and gender-related norms are not available, but could be created.

12. Are the relationships between task performance and neural signal(s) known?

i. The task assesses initial responsiveness to reward, operationalized in terms of neural response. Behaviorally, it is possible to examine win-stay/lose-shift strategies, though these data have related inconsistently to neural response to rewards.

13. Are the relationships between task performance and clinical feature(s) known?

i. Neural response to reward on the doors task has been related to depression and related constructs—both cross-sectionally and prospectively.

14. Is the task feasible for administration across sites?

i. Yes; it is currently being used at many research sites.

15. Can the task be used as a stand-alone behavioral task?

i. No

16. What work is needed to get this task ready for use in clinical trials?

i. Examine impact of pharmacological challenge
ii. Animal model of task
iii. Examine whether measures are sensitive to treatment response

17. Is the task copyrighted?

i. No.

References

1. Becker, M. P., Nitsch, A. M., Miltner, W. H., & Straube, T. (2014). A single-trial estimation of the feedback-related negativity and its relation to BOLD responses in a time-estimation task. The Journal of Neuroscience, 34(8), 3005-3012.
2. Bress, J. N., & Hajcak, G. (2013). Self-report and behavioral measures of reward sensitivity predict the feedback negativity. *Psychophysiology*, *50*(7), 610–616.
3. Bress, J. N., Meyer, A., & Proudfit, G. H. (2015). The stability of the feedback negativity and its relationship with depression during childhood and adolescence. *Development and Psychopathology*, *27*(4pt1), 1285–1294
4. Bress, J. N., Smith, E., Foti, D., Klein, D. N., & Hajcak, G. (2012). Neural response to reward and depressive symptoms in late childhood to early adolescence. *Biological Psychology*, *89*(1), 156–162.
5. Carlson, J. M., Foti, D., Mujica-Parodi, L. R., Harmon-Jones, E., & Hajcak, G. (2011). Ventral striatal and medial prefrontal BOLD activation is correlated with reward-related electrocortical activity: a combined ERP and fMRI study. Neuroimage, 57(4), 1608-1616.
6. Dichter, G. S., Felder, J. N., Petty, C., Bizzell, J., Ernst, M., & Smoski, M. J. (2009). The Effects of Psychotherapy on Neural Responses to Rewards in Major Depression. *Biological Psychiatry*, *66*(9), 886–897.
7. Forbes, E. E., Hariri, A. R., Martin, S. L., Silk, J. S., Moyles, D. L., Fisher, P. M., … Dahl, R. E. (2009). Altered Striatal Activation Predicting Real-World Positive Affect in Adolescent Major Depressive Disorder. *American Journal of Psychiatry*, *166*(1), 64–73.
8. Forbes, E. E., Ryan, N. D., Phillips, M. L., Manuck, S. B., Worthman, C. M., Moyles, D. L., Dahl, R. E. (2010). Healthy Adolescents' Neural Response to Reward: Associations With Puberty, Positive Affect, and Depressive Symptoms. *Journal of the American Academy of Child & Adolescent Psychiatry*, *49*(2), 162–172.
9. Foti, D., & Hajcak, G. (2009). Depression and reduced sensitivity to non-rewards versus rewards: Evidence from event-related potentials. *Biological Psychology*, *81*(1), 1–8.
10. Foti, D., & Hajcak, G. (2012). Genetic variation in dopamine moderates neural response during reward anticipation and delivery: Evidence from event-related potentials. *Psychophysiology*, *49*(5), 617–626.
11. Foti, D., Weinberg, A., Bernat, E. M., & Proudfit, G. H. (2015). Anterior cingulate activity to monetary loss and basal ganglia activity to monetary gain uniquely contribute to the feedback negativity. Clinical Neurophysiology, 126(7), 1338-1347.
12. Kujawa, A., Proudfit, G. H., & Klein, D. N. (2014). Neural reactivity to rewards and losses in offspring of mothers and fathers with histories of depressive and anxiety disorders. *Journal of Abnormal Psychology*, *123*(2), 287–297.
13. Levinson, A. R., Speed, B. C., & Hajcak, G. (2016). Reliability of the electrocortical response to gains and losses in the Doors task. *Under Review*.
14. Liu, W., Wang, L., Shang, H., Shen, Y., Li, Z., Cheung, E. F. C., & Chan, R. C. K. (2014). The influence of anhedonia on feedback negativity in major depressive disorder. *Neuropsychologia*, *53*, 213–220.
15. Luking, K. R., Nelson, B. D., Infantolino, Z. P., Sauder, C. L., & Hajcak, G. (under review). Internal reliability of fMRI and EEG Measures of Reward in Late Childhood and Early Adolescence.

16. Nelson, B. D., Perlman, G., Klein, D. N., Kotov, R., & Hajcak, G. (in press). Blunted Neural Response to Rewards Prospectively Predicts the Development of Depression in Adolescent Girls. *American Journal of Psychiatry*.
17. Parvaz, M. A., Gabbay, V., Malaker, P., & Goldstein, R. Z. (2016). Objective and specific tracking of anhedonia via event-related potentials in individuals with cocaine use disorders. *Drug and Alcohol Dependence*. doi:10.1016/j.drugalcdep.2016.05.004
18. Plichta, M. M., Schwarz, A. J., Grimm, O., Morgen, K., Mier, D., Haddad, L., Meyer-Lindenberg, A. (2012). Test–retest reliability of evoked BOLD signals from a cognitive–emotive fMRI test battery. *NeuroImage*, *60*(3), 1746–1758.
19. Speed, B. C., Nelson, B. D., Auerbach, R. P., Klein, D. N., & Hajcak, G. (2016). Depression Risk and Electrocortical Reactivity During Self-Referential Emotional Processing in 8 to 14 Year-Old Girls. *Journal of Abnormal Psychology*. doi:10.1037/abn0000173
20. Weinberg, A., & Shankman, S. A. (2016). Blunted Reward Processing in Remitted Melancholic Depression. *Clinical Psychological Science*. doi:10.1177/2167702616633158
21. Weinberg, A., Liu, H., Hajcak, G., & Shankman, S. A. (2015). Blunted neural response to rewards as a vulnerability factor for depression: Results from a family study. *Journal of Abnormal Psychology*, *124*(4), 878–889

Appendix PVS-IIb: Detailed Evaluation Criteria for Monetary Incentive Delay Task

PVS Construct: *Reward Responsiveness*

PVS Sub-construct: *Reward Anticipation*

1. **How valid a test of the construct is the task? Ok.**
   i. The MID task elicits robust and reliable brain activity (see below) during anticipation of monetary gains in the nucleus accumbens (NAcc) (3); during anticipation of monetary losses in the anterior insula, and less robust but still reliable activity in response to gain outcomes in the medial prefrontal cortex (MPFC) and putamen - possibly due to split trials) (5).
   ii. NAcc activity during anticipation of large gains often correlates with cue elicited positive arousal (7).
   iii. NAcc activity during anticipation of large gains also correlates selectively with individual differences in positive aroused traits (r's ~ .3) (4).
   iv. NAcc activity during anticipation of large gains correlates with negative symptoms across several studies of patients with schizophrenia (r's ~ .5), but not as robustly with symptoms related to affective disorders (10).

2. **Does the task have good psychometric characteristics (incl. high internal reliability, test-retest reliability, sensitivity/specificity, limited practice effects, availability of alternate forms, longitudinal stability)?** Yes.
   i. Internal reliability: Split–half reliability of neural activity during the first testing session indicated that right NAcc activity during anticipation of large gains (ICC = 0.56/0.71, p<.05) was moderate (4) (unpublished supplement).
   ii. Test-retest reliability: Test-retest reliability of neural activity over a > 2 year period indicated that peak right NAcc activity during anticipation of large gains (ICC=0.64/0.78) and right AIns activity during anticipation of large losses (ICC=.47/.64) was moderate to strong (4) (other conditions showed less significance and peaks showed better reliability than fitted contrasts). Similarly, other neuroimaging studies using comparable reward tasks demonstrated good reliability if they used large (6) but not small (8) incentives.
   iii. Measures with the greatest test-retest reliability were also the most correlated with affective traits (r~.3), while signal to noise ratio was not (4).
   iv. Power analysis indicated that for large effect sizes (f=3.07) typically observed in NAcc activity contrasts of anticipation of large versus no gains, 6 subjects were sufficient to detect a group effect at a power of .80 (p<.05).

v. Alternate forms (i.e., pseudorandom orders) are available and produce indistinguishable results.
vi. Developmental stability over adolescence is currently being assessed in large samples (e.g., IMAGEN).

3. **Are parameters for administering the task (e.g., number of trials, stimulus characteristics, etc.) standardized on an empirical basis?** Somewhat.
   i. Design: A recent and popular version of the MID task uses a 2 (valence: gain, loss) x 3 (magnitude: $0,$1,$5) factorial design with 15-18 trials per cell. Order is pseudorandom and balanced with a 2-6 second intertrial interval. Cue features can and have represented diverse alternative incentive features including probability (9), required effort (11) etc.
   ii. Analysis: Anticipation can and should be separately analyzed from outcomes (which are conditional on and orthogonal to anticipation). Raw averages of peak activation can be extracted and analyzed and show superior test-retest reliability to contrasts and resulting fits (4).
   iii. Development: Task parameters and requirements could benefit from continued optimization, particularly with respect to balancing task length against psychometric criteria (i.e., more research / funding is needed).

4. **Is the task free from floor/ceiling effects that could preclude use in subjects with a range of impairment?** Yes.
   i. The MID task avoids floor and ceiling effects by implementing an adaptive target response window that allows it to be administered in most subject populations, including clinical samples and across the lifespan (10)(14), which controls the expected value of cues and outcomes across diverse samples.

5. **Does the task have the same performance characteristics across cultures? Is it free from culture-and language-specific features/stimuli?** Yes.
   i. The MID task has been applied across diverse cultures where fMRI is available and produced qualitatively comparable results (e.g., Britain, Germany, France, Netherlands, Israel, Japan, China, US, etc.) (12). Training subjects with abstract cues helps control pre-existing confounds due to learning or pre-existing symbolic associations. Cue mappings can also be fully counterbalanced within datasets.
   ii. Abstract cues facilitate mapping incentives according to culturally equivalent incentive schemes (e.g., adapting the symbol $ to €) to be determined by culturally-informed researchers.

6. **Is the task sensitive to change and lack and loss of function?** Yes.
   i. Greater age-related declines in AIns activity during anticipation of large losses versus NAcc activity during anticipation of large gains has been replicated in multiple studies (14; 15).
   ii. Some clinical research suggests that blunted NAcc activity during gain anticipation in schizophrenic patients on typical antipsychotics can partially be reversed after switching to atypical antipsychotics, in tandem with diminution of negative symptoms (13).

7. **Can the task (or its analog) be used in animals?** Yes (after substituting primary for secondary rewards).
   i. Tasks that vary cued reward magnitude (i.e., drops of sugar water) elicit magnitude-dependent increases in NAcc dopamine release in rats, as assessed by in vivo cyclic voltammetry (16) (the same is not true for cued effort, paralleling human studies).

8. **Can the task be used across age groups?** Yes.
   i. The MID task has been used in adolescents and elders, and tokenized versions have been extended to children (but require norming).
   ii. Adolescents (<18) show qualitatively similar activity patterns, with somewhat diminished NAcc activity during anticipation of large gains (17; 18).
   iii. Older adults (>60) typically show similar activity as younger adults, with the exception of less AIns activity during anticipation of large losses (4; 14)

9. **Can the task be used with methods to interrogate brain circuitry (e.g., fMRI, EEG, etc.)?** Yes (particularly FMRI).
   i. The MID task is specifically designed and optimized for use with FMRI.
   ii. The MID task has been used with EEG, but deep sources are difficult to localize (6).
   iii. We are currently exploring connections with raclopride displacement PET but disparate timescales are difficult to compare (but see (19)).

10. **Is there consensus on which metric/score should be considered to be primary?** Somewhat.
   i. While there is no formal consensus for neural activity, the majority of researchers contrast: (1) gain versus nongain anticipation; (2) loss versus nonloss anticipation; (3) gain versus nongain outcome; (4) nonloss versus loss outcome. Alternatively, researchers extract peak activation for all conditions (e.g., valence by magnitude) from volumes of interest in the NAcc, MFPC, and right AIns (recommended). (12)
   ii. Valence and arousal ratings for each of the incentive cues can be collected after (or even during) the task, mean-deviated, and rotated to derive cue elicited positive arousal and negative arousal scores (20).

iii. Functional connectivity between nodes could be extracted for specific trial phases and conditions, but these indices have not received extensive psychometric characterization (e.g.,(21)).

11. Are adequate normative data available across age, gender, education, ethnicity, SES? Partially.

i. Some normative data are available in medium-sized samples (n=52) for age and gender (4), and larger datasets are coming online (e.g., IMAGEN). Samples to date have tended to include high education and socioeconomic status individuals (except in cases of clinical groups).

12. Are the relationships between task performance and neural signal(s) known? Yes.

i. Behavioral performance is typically controlled so that associations between overt behavior and brain activity are dissociable.

ii. Regressors that parametrically model reaction time in response to each target, however, typically robustly activate the putamen and supplementary motor cortex (22).

13. Are the relationships between task performance and clinical feature(s) known? Somewhat.

i. Behavioral performance is typically controlled so that associations between overt behavior and brain activity are dissociable (as above).

ii. The strongest clinical correlates of NAcc activity during gain anticipation to date have included negative symptoms in the context of schizophrenia (23) and hyperactive symptoms in the context of ADHD (6; 24).

iii. Many other disorders remain to be explored (e.g., affective disorders, addiction).

14. Is the task feasible for administration across sites? Yes.

i. An adapted version of the MID task has been used in approximately 2000 youth across 8 European sites in the IMAGEN consortium, and is also being used in another multisite study (FAST-MAS). Initial verification of adequate signal homogeneity and spatiotemporal resolution across scanners is essential.

15. Can the task be used as a stand-alone behavioral task? Possibly not.

i. Because the MID task adaptively controls performance to equate expected value, faster reaction time measures to provide limited information.

ii. Researchers can, however, solicit affective responses (typically valence and arousal) to incentive cues as a summary measure of affective responsiveness (9). Combination of neural self-report measures, however, is recommended.

16. What work is needed to get this task ready for use in clinical trials? In use, but more could be done.

i. The current "standard" 3 (magnitude) x 2 (valence) version is already in use in many clinical and pharmacological protocols (however, see below):

ii. The task involves a speeded reaction time response. This is controlled across incentive conditions, but may add to the observed signal. If reduced motor engagement is desired, a MID task version involving choices rather than speeded reaction time could be compared with canonical versions (this would require piloting, however, since it could change the affective responses and generalizability of the task).

iii. More extensive sets of gain and loss magnitudes could be investigated in a longer experiment to determine optimal magnitudes (however, set effects may also play a role).

iv. A directly parallel version could be devised and characterized in rats using both older (voltammetry) and newer (optogenetic fiber photometry) measures, possibly alongside pharmacological modulation for validation.

v. Faster peripheral physiological measures (facial electromyography + pupillary dilation) might be tested as a potentially diluted but implicit behavioral probe of affective responses during the MID task.

vi. Task parameters (i.e., number and composition of conditions, number of trials per condition, minimum variable intertrial interval) could be compared and optimized in a series of trials.

17. Is the task copyrighted? No.

i. Initial development of the MID Task was funded through an NIH B/START grant MH066923 so the task belongs to American taxpayers. Version control, however, is maintained by BK (knutson@stanford.edu), who can provide recent copies of the task upon request.

References

1. Knutson B, Westdorp a, Kaiser E, Hommer D (2000): FMRI visualization of brain activity during a monetary incentive delay task. Neuroimage 12: 20–7.
2. Haber SN, Knutson B (2010): The reward circuit: linking primate anatomy and human imaging. Neuropsychopharmacology 35: 4–26.

3. Knutson B, Adams CM, Fong GW, Hommer D (2001): Anticipation of increasing monetary reward selectively recruits nucleus accumbens. . . J Neurosci (Vol. 21) . doi: 20015472 [pii].
4. Wu CC, Samanez-Larkin GR, Katovich K, Knutson B (2014): Affective traits link to reliable neural markers of incentive anticipation. Neuroimage 84: Elsevier Inc.279–289.
5. Knutson B, Fong GW, Bennett SM, Adams CM, Hommer D (2003): A region of mesial prefrontal cortex tracks monetarily rewarding outcomes: characterization with rapid event-related fMRI. Neuroimage 18: 263–272.
6. Plichta MM, Scheres A (2014): Ventral-striatal responsiveness during reward anticipation in ADHD and its relation to trait impulsivity in the healthy population: A meta-analytic review of the fMRI literature. Neurosci Biobehav Rev 38: 125–134.
7. Knutson B, Greer SM (2008): Anticipatory affect: Neural correlates and consequences for choice. Philos Trans R Soc Lond B Biol Sci 363: 3771–86.
8. Fliessbach K, Rohe T, Linder NS, Trautner P, Elger CE, Weber B (2010): Retest reliability of reward- related BOLD signals. Neuroimage 50: Elsevier Inc.1168–1176.
9. Knutson B, Taylor J, Kaufman M, Peterson R, Glover G (2005): Distributed neural representation of expected value. J Neurosci 25: 4806–4812.
10. Knutson B, Heinz A (2015): Probing psychiatric symptoms with the monetary incentive delay task. Biol Psychiatry 77: 418–20.
11. Croxson PL, Walton ME, O'Reilly JX, Behrens TEJ, Rushworth MFS (2009): Effort-based cost-benefit valuation and the human brain. J Neurosci 29: 4531–41.
12. Balodis IM, Potenza MN (2015): Anticipatory reward processing in addicted populations: A focus on the monetary incentive delay task. Biol Psychiatry 77: Elsevier434–444.
13. Schlagenhauf F, Juckel G, Koslowski M, Kahnt T, Knutson B, Dembler T, et al. (2008): Reward system activation in schizophrenic patients switched from typical neuroleptics to olanzapine. Psychopharmacology (Berl) 196: 673–684.
14. Samanez-Larkin GR, Gibbs SEB, Khanna K, Nielsen L, Carstensen LL, Knutson B (2007): Anticipation of monetary gain but not loss in healthy older adults. Nat Neurosci 10: 787–791.
15. Samanez-Larkin GR, Levens SM, Perry LM, Dougherty RF, Knutson B (2012): Frontostriatal white matter integrity mediates adult age differences in probabilistic reward learning. J Neurosci 32: 5333–7.
16. Gan JO, Walton ME, Phillips PEM (2010): Dissociable cost and benefit encoding of future rewards by mesolimbic dopamine. Nat Neurosci 13: 25–7.
17. Bjork JM, Knutson B, Fong GW, Caggiano DM, Bennett SM, Hommer DW (2004): Incentive-elicited brain activation in adolescents: similarities and differences from young adults. J Neurosci 24: 1793– 1802.
18. Bjork JM, Knutson B, Hommer DW (2008): Incentive-elicited striatal activation in adolescent children of alcoholics. Addiction 103: 1308–1319.
19. Schott BH, Minuzzi L, Krebs RM, Elmenhorst D, Lang M, Winz OH, et al. (2008): Mesolimbic functional magnetic resonance imaging activations during reward anticipation correlate with reward-related ventral striatal dopamine release. J Neurosci 28: 14311–9.

20. Knutson B, Katovich K, Suri G (2014): Inferring affect from fMRI data. Trends Cogn Sci 18: 422–8.
21. Cho YT, Ernst M, Fudge JL (2013): Cortico-amygdala-striatal circuits are organized as hierarchical subsystems through the primate amygdala. J Neurosci 33: 14017–30.
22. Knutson B, Taylor J, Kaufman M, Peterson R, Glover G (2005): Distributed neural representation of expected value. J Neurosci 25: 4806–12.
23. Juckel G, Schlagenhauf F, Koslowski M, Wüstenberg T, Villringer A, Knutson B, et al. (2006): Dysfunction of ventral striatal reward prediction in schizophrenia. Neuroimage 29: 409–16.
24. Scheres A, Milham MP, Knutson B, Castellanos FX (2007): Ventral striatal hyporesponsiveness during reward anticipation in attention-deficit/hyperactivity disorder. Biol Psychiatry 61: 720–724.

Appendix PVS-IIc: Detailed Evaluation Criteria for Devaluation Learning Tasks

PVS Construct: *Reward Learning*

PVS Sub-construct: *Habit*

1. How valid a test of the construct is the task?

Devaluation learning tasks (DLT) were developed to provide an objective measure of participants' ability to establish a habitual response to stimuli associated with outcomes that earn points, and then a 'slips-of-action' phase measures their ability to not respond to formerly rewarded stimuli that are devalued in the final test phase. Prior studies have shown that devaluation learning:

1) is negatively correlated with tic severity in patients with Gilles de la Tourette Syndrome (Delorme et al. 2016), OCD traits (Snorrason et al. 2016) in young adults, and blunted in patients with OCD (Gillan et al. 2011) and patients with alcohol dependence (Sjoerds et al. 2013). Medicated patients with Gilles de la Tourette Syndrome showed improved devaluation learning over non-medicated patients (Delorme et al. 2016). Patients with longer duration of alcohol dependence showed less engagement of the ventromedial prefrontal cortex (a critical structure for goal-directed behavior that can override habitual behavior) (Sjoerds et al. 2013).
2) is blunted with acute dopamine (de Wit et al. 2012) and tryptophan depletion (Worbe et al. 2015), and steeper declines in plasma tryptophan levels predicted poorer performance in devaluing stimuli (Worbe et al. 2015)
3) is linked to reward-related activation(de Wit et al. 2009) in humans and rats (Smith & Graybiel 2016) (dorsal striatum) and atypical structural connectivity from reward nodes to motor regions (Delorme et al. 2016) in patients with Gilles de la Tourette Syndrome.

2. Does the task have good psychometric characteristics (incl. high internal reliability, test-retest reliability, sensitivity/specificity, limited practice effects, availability of alternate forms, longitudinal stability)?

i. high internal reliability: not evaluated
ii. test-retest reliability: not evaluated
iii. Sensitivity/specificity: not evaluated
iv. Limited practice effects: not evaluated
v. Availability of alternate forms: not evaluated but possible
vi. Longitudinal stability: not evaluated

3. **Are parameters for administering the task (e.g., number of trials, stimulus characteristics, etc.) standardized on an empirical basis?**
   i. No. Parameters need to be optimized for administration across adult, pediatric, and clinical populations.

4. **Is the task free from floor/ceiling effects which would preclude use in subjects with a range of impairment?**
   i. Incomplete; task characteristics show adequate variability across healthy adult and adult clinical populations, but additional work is needed for pediatric populations.

5. **Does the task have the same performance characteristics across cultures? Is it free from culture- and language-specific features/stimuli?**
   i. Not yet known

6. **Is the task sensitive to change and lack and loss of function?**
   i. Yes, as evidenced by psychopharmacologic challenge (Worbe et al. 2015; de Wit et al. 2012) and findings in patient populations (Sjoerds et al. 2013; Gillan et al. 2011; Delorme et al. 2016).

7. **Can the task (or its analog) be used in animals?**
   i. Incomplete – one study to date shows effective animal analog (Smith & Graybiel 2016).

8. **Can the task be used across age groups?**
   i. Versions implemented in adults (Snorrason et al. 2016; Sjoerds et al. 2013; Delorme et al. 2016; Gillan et al. 2011; de Wit et al. 2012; Worbe et al. 2015), older adults (de Wit et al. 2011), and children (Geurts & de Wit 2013), but the childhood task did not elicit the intended devaluation effect in children with autism. Unclear if this is because habit formation is intact in children with autism or the task was not properly optimized to be sensitive to differences in performance for children—existing data on reward systems (Kohls et al. 2012; Dichter et al. 2012) and reversal learning in autism (Yerys et al. 2009; D'Cruz et al. 2013; Reed et al. 2011) would suggest devaluation learning to be a reasonable target to expect differences between groups.

9. **Can the task be used with methods to interrogate brain circuitry (e.g., fMRI, EEG, etc.)?**
   i. Yes (see above). The task has been implemented with fMRI and DTI (de Wit et al. 2009; Delorme et al. 2016)

10. **Is there consensus on which metric/score should be considered to be primary?**

i. Difference score of valuable minus devalued response % in Slips-of-Action and Baseline phases
ii. Correct/Incorrect responses and decreases in RT over the course of the learning phase
iii. Accuracy during the Outcome-devaluation stage

11. Are adequate normative data available across age, gender, education, ethnicity, SES?

i. No

12. Are the relationships between task performance and neural signal(s) known?

i. Yes (see above). Linkage to striatum in humans (Delorme et al. 2016; de Wit et al. 2009) and animals (Smith & Graybiel 2016).

13. Are the relationships between task performance and clinical feature(s) known?

i. Yes, correlates with tic severity in Gilles de la Tourette Syndrome (Delorme et al. 2016) and poorer devaluation observed in patients with OCD (Gillan et al. 2011) and alcohol dependence (Sjoerds et al. 2013) but not Parkinson's (de Wit et al. 2011) or autism (Geurts & de Wit 2013).

14. Is the task feasible for administration across sites?

i. Yes

15. Can the task be used as a stand-alone behavioral task?

i. Yes

16. What work is needed to get this task ready for use in clinical trials?

i. Optimization across pediatric and clinical pediatric populations, validity and normalization analyses.

17. Is the task copyrighted?

i. No

References:

1. Delorme, C. *et al.* Enhanced habit formation in Gilles de la Tourette syndrome. *Brain* **139,** 605–615 (2016).
2. Snorrason, I., Lee, H. J., de Wit, S. & Woods, D. W. Are nonclinical obsessive-compulsive symptoms associated with bias toward habits? *Psychiatry Res.* **241,** 221–223 (2016).
3. Gillan, C. M. *et al.* Disruption in the balance between goal-directed behavior and habit learning in obsessive-compulsive disorder. *Am. J. Psychiatry* **168,** 718–726 (2011).

4. Sjoerds, Z. *et al.* Behavioral and neuroimaging evidence for overreliance on habit learning in alcohol-dependent patients. *Transl. Psychiatry* **3,** e337 (2013).
5. de Wit, S. *et al.* Reliance on habits at the expense of goal-directed control following dopamine precursor depletion. *Psychopharmacology (Berl.)* **219,** 621–631 (2012).
6. Worbe, Y., Savulich, G., Wit, S. de, Fernandez-Egea, E. & Robbins, T. W. Tryptophan Depletion Promotes Habitual over Goal-Directed Control of Appetitive Responding in Humans. *Int. J. Neuropsychopharmacol.* **18,** pyv013 (2015).
7. de Wit, S., Corlett, P. R., Aitken, M. R., Dickinson, A. & Fletcher, P. C. Differential Engagement of the Ventromedial Prefrontal Cortex by Goal-Directed and Habitual Behavior toward Food Pictures in Humans. *J. Neurosci.* **29,** 11330–11338 (2009).
8. Smith, K. S. & Graybiel, A. M. Habit formation coincides with shifts in reinforcement representations in the sensorimotor striatum. *J. Neurophysiol.* **115,** 1487–1498 (2016).
9. de Wit, S., Barker, R. A., Dickinson, A. D. & Cools, R. Habitual versus Goal-directed Action Control in Parkinson Disease. *J. Cogn. Neurosci.* **23,** 1218–1229 (2011).
10. Geurts, H. M. & de Wit, S. Goal-directed action control in children with autism spectrum disorders. *Autism* **18,** 409–18 (2013).
11. Kohls, G., Chevallier, C., Troiani, V. & Schultz, R. T. Social 'wanting' dysfunction in autism: neurobiological underpinnings and treatment implications. *J. Neurodev. Disord.* **4,** 10 (2012).
12. Dichter, G. S., Damiano, C. A. & Allen, J. A. Reward circuitry dysfunction in psychiatric and neurodevelopmental disorders and genetic syndromes: animal models and clinical findings. *J. Neurodev. Disord.* **4,** 19 (2012).
13. Yerys, B. E. *et al.* Set-shifting in children with autism spectrum disorders: reversal shifting deficits on the Intradimensional/Extradimensional Shift Test correlate with repetitive behaviors. *Autism* **13,** 523–538 (2009).
14. D'Cruz, A.-M. *et al.* Reduced behavioral flexibility in autism spectrum disorders. *Neuropsychology* **27,** 152–160 (2013).
15. Reed, P., Watts, H. & Truzoli, R. Flexibility in young people with autism spectrum disorders on a card sort task. *Autism* (2011). doi:10.1177/1362361311409599

Appendix PVS-IId: Detailed Evaluation Criteria for Probabilistic Reward Task

PVS Construct: *Reward Learning*

PVS Sub-construct: *Probabilistic and Reinforcement Learning*

1. How valid a test of the construct is the task?

The Probabilistic Reward task (PRT) was developed to provide an objective measure of participants' ability to modify behavior as a function of reward (Pizzagalli et al. 2005) (modified after (Tripp & Alsop 1999)), and yields measures of reward responsiveness and reward learning. Prior studies have shown that response bias towards a more frequently rewarded stimulus:

(1) is inversely related to current anhedonic symptoms in unselected adults, individuals with elevated depressive symptoms, and unmedicated individuals with current MDD (Pizzagalli, Iosifescu, et al. 2008; Pizzagalli et al. 2005; Bogdan & Pizzagalli 2006) and relatives of patients with major depression (W.-H. Liu et al. 2016); correlates with reduced hedonic capacity/approach motivation prepubertal children (Luking et al. 2015); and correlates with cigarette craving among smokers (Peechatka et al. 2015);
(2) predicts self-reported anhedonic symptoms 38 days later (Pizzagalli et al. 2005);
(3) is blunted in individuals with increased depressive symptoms, current MDD, and past MDD (Pizzagalli, Iosifescu, et al. 2008; Pizzagalli et al. 2005; Pechtel et al. 2013; Whitton et al. 2016; Liu et al. 2011), particularly those with elevated anhedonic symptoms (Vrieze, Pizzagalli, et al. 2013) or melancholic depression (Fletcher et al. 2015); in relatives of patients with major depression with sub-clinical depressive symptoms (W.-H. Liu et al. 2016); and in youth reporting anhedonia across various DSM diagnosis (Morris et al. 2015).
(4) is improved by pharmacological treatments among depressed inpatients (Vrieze, Pizzagalli, et al. 2013) and by residential treatment in female adolescents with co-occurring depression and substance abuse (Boger et al. 2014);
(5) is linked to both resting (Webb et al. 2016) (Kaiser et al., under review) and reward-related activation and functional connectivity within nodes of the brain reward system (ventral/dorsal striatum, orbitofrontal cortex, dorsal anterior cingulate cortex) (Santesso et al. 2008a; Santesso et al. 2009; Bogdan et al. 2011) as well as ERP markers of reinforcement learning (Santesso et al. 2008a; Whitton et al. 2016; Bress & Hajcak 2013);
(6) is linked to DA release in extrastriatal regions (as assessed by PET) (Vrieze, Ceccarini, et al. 2013);
(7) is associated with genetic variants known to modulate prefrontal dopaminergic variation (*COMT*; (Lancaster et al. 2012; Lancaster et al. 2015; Goetz et al. 2013); and (Corral-Frías et al. 2016), risk for mood disorders and schizophrenia (*CACNA1C*; (Lancaster et al. 2014), and mu-opioid receptor function (Lee et al. 2011).

(8) is potentiated or blunted by pharmacological challenges hypothesized to increase (e.g., nicotine, amphetamine) or decrease (e.g., single low doses of pramipexole thought to reduce phasic DA response via presynaptic autoreceptor activation), respectively, DA signaling in both humans and rats (Barr et al. 2008; Pizzagalli, Evins, et al. 2008; Pergadia et al. 2014; Der-Avakian et al. 2013);
(9) is potentiated in healthy controls by high-frequency rapid TMS over the left dorsolateral prefrontal cortex (Ahn et al. 2013);
(10) is blunted by acute laboratory and prolonged naturalistic stressors (Bogdan & Pizzagalli 2006; Pizzagalli et al. 2007), particularly in individuals carrying genetic variants previously associated with increased HPA reactivity or depression (Bogdan et al. 2011; Bogdan et al. 2010; Nikolova et al. 2012); conversely, a greater response bias under stress among individuals with General Anxiety Disorders (GAD) predicts lower depression symptoms 1 month later (Morris & Rottenberg 2015).
(11) is heritable (46%) (Bogdan & Pizzagalli 2009).

2. **Does the task have good psychometric characteristics (incl. high internal reliability, test-retest reliability, sensitivity/specificity, limited practice effects, availability of alternate forms, longitudinal stability)?**
   i. high internal reliability: not evaluated
   ii. test-retest reliability: 0.57 over 38 days in unselected individuals (Pizzagalli et al. 2005); replicated in an independent unselected sample: r = 0.50-0.56 over 39 days (Santesso et al. 2008b)
   iii. Sensitivity/specificity: ROC analyses not performed yet (but data are available)
   iv. Limited practice effects: minimized by using different alternate forms (see below).
   v. Availability of alternate forms: Yes (5 forms)
   vi. Longitudinal stability: limited (only evaluated over ~40 days)

3. **Are parameters for administering the task (e.g., number of trials, stimulus characteristics, etc.) standardized on an empirical basis?**
   i. Task characteristics (e.g., stimulus size and exposure) were optimized in order to achieve an overall accuracy ~0.85 (to allow condition or group modulations).

4. **Is the task free from floor/ceiling effects which would preclude use in subjects with a range of impairment?**
   i. Yes; task characteristics (e.g., stimulus size and exposure) were optimized in order to achieve an overall accuracy ~0.85.

5. **Does the task have the same performance characteristics across cultures? Is it free from culture- and language-specific features/stimuli?**

i. Findings of reduced response bias in MDD vs. healthy controls have been replicated across US (Pizzagalli, Iosifescu, et al. 2008), European (Vrieze, Pizzagalli, et al. 2013), and Asian (Liu et al. 2011) samples. The task and its instructions have been translated in several languages (e.g., Dutch, English, German, Korean, Chinese)
ii. The task has been freely disseminated by >110 research groups across many countries; its minimal reliance on verbal stimuli makes cross-cultural comparisons feasible.

6. Is the task sensitive to change and lack and loss of function?

i. Yes, as evidenced by pharmacological challenges (Barr et al. 2008; Pizzagalli, Evins, et al. 2008; Pergadia et al. 2014), neurostimulation (Ahn et al. 2013), and findings in patient samples (Pizzagalli, Iosifescu, et al. 2008; Fletcher et al. 2015; Vrieze, Pizzagalli, et al. 2013; Liu et al. 2011).

7. Can the task (or its analog) be used in animals?

i. Yes; a conceptually analogous version has been developed for rats (Der-Avakian et al. 2013). Cross-species studies have shown that the same findings have emerged in humans and rats when using pharmacological challenges (Pizzagalli, Evins, et al. 2008; Der-Avakian et al. 2013), nicotine withdrawal (Pergadia et al. 2014), or stressors (Bogdan & Pizzagalli 2006) (and Der-Avakian et al., in preparation).

8. Can the task be used across age groups?

i. Some (unpublished) data in children

9. Can the task be used with methods to interrogate brain circuitry (e.g., fMRI, EEG, etc.)?

i. Yes (see above). The task has been used in conjunction with ERP, fMRI and PET.

10. Is there consensus on which metric/score should be considered to be primary?

i. Response bias
ii. Reward learning (e.g., RB(block 3) – RB(Block 1)
iii. Secondary: Discriminability, accuracy, RT

11. Are adequate normative data available across age, gender, education, ethnicity, SES?

i. The task has been administered to over 1,000 individuals. Age- and gender-related norms are available.

12. Are the relationships between task performance and neural signal(s) known?

i. Yes (see point #1)

13. Are the relationships between task performance and clinical feature(s) known?

i. Yes (see point #1)

14. Is the task feasible for administration across sites?

i. Yes (standardization and a manual have been developed for the EMBARC study).
ii. In addition, Dr. Pizzagalli's lab has feely provided the task to over 110 groups since 2005, and extensive documentation/manuals are available for standardization across sites.

15. Can the task be used as a stand-alone behavioral task?

i. Yes

16. What work is needed to get this task ready for use in clinical trials?

i. EMBARC and CNTRACS will provide info regarding its use in clinical trials

17. Is the task copyrighted?

i. The code is copyrighted (Harvard University) and is provided for free to academic groups. A license agreement is required for industry.

References:

See overall bibliography

Appendix PVS-IIe: Detailed Evaluation Criteria for Risk and Ambiguity Task

PVS Construct: *Reward Valuation*

PVS Sub-construct: *Reward (probability)*

1. How valid a test of the construct is the task?

The Risk and Ambiguity Task was developed to assess individual behavior under uncertainty. It assesses individual attitudes towards risk (known outcome probabilities) and ambiguity (unknown outcome probabilities)[1,2]. The task yields measures of risk and ambiguity attitudes in the gain and loss domains, as well as measures of decision quality[3]. Prior studies have shown:

1) Decreased ambiguity aversion in adolescents compared to adults[4], which increases with age[5]. No ambiguity aversion in pre-adolescent children[6].
2) Increased risk aversion in the gain domain and risk seeking in the loss domain, as well as decreased decision quality, in older adults[3].
3) Increased aversion to ambiguous losses in individuals with PTSD, which is correlated with symptom strength and mediates the association between the degree of combat exposure and the degree of symptoms, specifically anxious arousal[7].
4) Decreased decision quality and increased ambiguity aversion in individuals with OCD[8].
5) Correlation between the gray-matter volume of a region in right Posterior Parietal Cortex and individual risk tolerance[9].
6) Effect of individual risk and ambiguity attitudes on activation magnitude in value-related brain areas[1].

2. Does the task have good psychometric characteristics (incl. high internal reliability, test-retest reliability, sensitivity/specificity, limited practice effects, availability of alternate forms, longitudinal stability)?

i. high internal reliability: not evaluated
ii. test-retest reliability: risk and ambiguity attitudes stable in 18 subjects across two sessions separated by several days[1], but reliability not quantified.
iii. Sensitivity/specificity: not evaluated
iv. Limited practice effects: not evaluated
v. Availability of alternate forms: several versions of the task have been used
vi. Longitudinal stability: not evaluated

3. Are parameters for administering the task (e.g., number of trials, stimulus characteristics, etc.) standardized on an empirical basis?

i. No

4. **Is the task free from floor/ceiling effects which would preclude use in subjects with a range of impairment?**
   i. The task has been used in a wide range of ages, including several psychiatric conditions (see above), with no floor/ceiling effects.

5. **Does the task have the same performance characteristics across cultures? Is it free from culture- and language-specific features/stimuli?**
   i. The task is free from culture- and language-specific stimuli. Performance across cultures not evaluated.

6. **Is the task sensitive to change and lack and loss of function?**
   i. Not evaluated

7. **Can the task (or its analog) be used in animals?**
   i. In principle the task is specific to humans, as it requires understanding of symbolic probabilities. One study, however, used an analog in monkeys[10].

8. **Can the task be used across age groups?**
   i. Yes, the task has been successfully used across a wide range of ages, from 8 to 90 years old[3-6].

9. **Can the task be used with methods to interrogate brain circuitry (e.g., fMRI, EEG, etc.)?**
   i. Yes, the task has been used with fMRI[1].

10. **Is there consensus on which metric/score should be considered to be primary?**
    i. Risk and ambiguity attitudes in the gain and loss domains, estimated based on proportion of choices of each type, or based on fitting a behavioral model
    i. Estimates of decision quality based on violations of first-order stochastic dominance and transitivity

11. **Are adequate normative data available across age, gender, education, ethnicity, SES?**
    i. No

12. **Are the relationships between task performance and neural signal(s) known?**
    i. Only partially (see above)

13. Are the relationships between task performance and clinical feature(s) known?

i. Relationships with some features, including obsessive compulsive disorder and anxious arousal are known (see above)

14. Is the task feasible for administration across sites?

i. Yes. It includes standardized instructions and has been used in a multi-site study at Yale and NYU[3,4]. The task was also provided upon request to several other groups.

15. Can the task be used as a stand-alone behavioral task?

i. Yes.

16. What work is needed to get this task ready for use in clinical trials?

i. Standardization of parameters, estimation of the minimal task length required for adequate parameter estimates.

17. Is the task copyrighted?

i. No

References

1. Levy I, Snell J, Nelson AJ, Rustichini A, Glimcher PW. Neural representation of subjective value under risk and ambiguity. *J Neurophysiol.* 2010;103(2):1036-1047.
2. Levy I, Rosenberg Belmaker L, Manson K, Tymula A, Glimcher PW. Measuring the subjective value of risky and ambiguous options using experimental economics and functional MRI methods. *J Vis Exp.* 2012(67):e3724.
3. Tymula A, Rosenberg Belmaker LA, Ruderman L, Glimcher PW, Levy I. Like cognitive function, decision making across the life span shows profound age-related changes. *Proc Natl Acad Sci U S A.* 2013;110(42):17143-17148.
4. Tymula A, Rosenberg Belmaker LA, Roy AK, et al. Adolescents' risk-taking behavior is driven by tolerance to ambiguity. *Proc Natl Acad Sci U S A.* 2012;109(42):17135-17140.
5. Blankenstein NE, Crone EA, van den Bos W, van Duijvenvoorde AC. Dealing With Uncertainty: Testing Risk- and Ambiguity-Attitude Across Adolescence. *Dev Neuropsychol.* 2016:1-16.
6. Li R, Brannon EM, Huettel SA. Children do not exhibit ambiguity aversion despite intact familiarity bias. *Front Psychol.* 2014;5:1519.
7. Ruderman L, Ehrlich DB, Roy A, Pietrzak RH, Harpaz-Rotem I, Levy I. Posttraumatic Stress Symptoms and Aversion to Ambiguous Losses in Combat Veterans. *Depress Anxiety.* 2016.
8. Pushkarskaya H, Tolin D, Ruderman L, et al. Decision-making under uncertainty in obsessive-compulsive disorder. *J Psychiatr Res.* 2015;69:166-173.

9. Gilaie-Dotan S, Tymula A, Cooper N, Kable JW, Glimcher PW, Levy I. Neuroanatomy predicts individual risk attitudes. *J Neurosci.* 2014;34(37):12394-12401.
10. Hayden BY, Heilbronner SR, Platt ML. Ambiguity aversion in rhesus macaques. *Front Neurosci.* 2010;4.

Appendix PVS-IIf: Detailed Evaluation Criteria for the Effort-Expenditure for Rewards Task

PVS Construct: *Reward Valuation*

PVS Sub-construct: *Effort*

1. How valid a test of the construct is the task?

The Effort Expenditure for Rewards Task (EEfRT; pronounced "Effort") was developed as a homologue to well-studied effort-based decision-making tasks used in the rodent literature (e.g., [1]). The task assesses an individual preference between expending greater physical effort in the form of speeded button presses in order to gain larger monetary rewards vs. less effort for smaller rewards. The task takes approximately 20 minutes, during which time individuals typically complete approximately 50 trials for which they must choose between "hard tasks" and "easy tasks". Trials vary in terms of both the magnitude of reward available for choosing the more effortful option as well as the probability of reward receipt. These manipulations were included to increase the ecological validity of the task for use in humans, but their inclusion means that the task is not a "pure" measure of effort alone.

(1) Consistent with predictions from the animal literature regarding the role of dopamine (DA) in effort-based decision-making [2], administration of the DA-releasing agent d-amphetamine increases the proportion of hard-tasks selected on the EEfRT [3]. Additionally the proportion of hard-tasks predicts amphetamine-induced DA release in the striatum [4].

(2) Proportion of hard-tasks is inversely related to trait anhedonia in an undergraduate sample enriched for anhedonia [5]; and positively related to trait reward anticipation and behavioral activation [6] (and unpublished observations);

(3) Repeated studies of patients in depression [7, 8] and schizophrenia [9-13]; have found evidence for altered performance on the EEfRT as compared to healthy controls. Importantly however, it remains unclear whether both groups show a primary reduction in effort expenditure (preference for less effortful options) or a deficit in effort allocation (exerting greater effort for trials that offer relatively smaller rewards, and failing to expend effort for trials that offer greater rewards). Further within these groups, relationships between EEfRT performance and dimensional measures of anhedonia or negative symptoms have been moderate at best, and often inconsistent, suggesting that they may be assessing different aspects of reward-related symptoms.

2. Does the task have good psychometric characteristics (incl. high internal reliability, test-retest reliability, sensitivity/specificity, limited practice effects, availability of alternate forms, longitudinal stability)?

i. high internal reliability: evaluation in process; will be available in winter 2016

ii. test-retest reliability: assessed in a sample of schizophrenia patients only. ICCs at 4 weeks range in this sample varied for different parameters (reward magnitude; probability). Range from .079 to 0.53.
iii. Sensitivity/specificity: unknown
iv. Limited practice effects: repeated assessments have not found significant evidence of practice effects [3, 14].
v. Availability of alternate forms: Yes – a wide number of variants are currently in use
vi. Longitudinal stability: limited (only evaluated for ~30 days)

3. Are parameters for administering the task (e.g., number of trials, stimulus characteristics, etc.) standardized on an empirical basis?

i. Partially. Task parameters are set to facilitate high completion rates (>90%) of high and low effort option in healthy or mild/moderate psychiatric patients.

4. Is the task free from floor/ceiling effects which would preclude use in subjects with a range of impairment?

i. Partially. Some subjects will make all of one choice type (i.e., all easy or all hard choices). This is estimated to occur <5% based in healthy participants, but may be more prevalent in patient populations with varying impairments.

5. Does the task have the same performance characteristics across cultures? Is it free from culture- and language-specific features/stimuli?

i. Findings in the US [8] have been replicated in an independently-collected Chinese sample [7].
ii. The task has been shared >85 research groups across 7 countries. Modification of reward magnitudes may be necessary to adjust for differences in currency valuations.

6. Is the task sensitive to change and lack and loss of function?

i. Repeated-administrations have shown sensitivity of the task to manipulations of dopamine (amphetamine)[3] and adenosine [3].

7. Can the task (or its analog) be used in animals?

i. Yes – the task was modeled after well-known effort-based decision-making paradigms in rodents[1, 15, 16]

8. Can the task be used across age groups?

i. Some (unpublished) data have been administered in children.

9. Can the task be used with methods to interrogate brain circuitry (e.g., fMRI, EEG, etc.)?

i. Yes – published data have linked the task of fMRI [17], EEG [4, 18], and dopamine-receptor PET imaging [4]

10. Is there consensus on which metric/score should be considered to be primary?

i. Total proportion of hard task choices
ii. Difference in proportion of hard task choices for high vs. low probability levels and/or reward

11. Are adequate normative data available across age, gender, education, ethnicity, SES?

i. The task has been administered in over 1,000 individuals. Age, Sex, IQ and SES norms are in the process of being compiled, and are expected to be published in the winter of 2016.

12. Are the relationships between task performance and neural signal(s) known?

i. Partially – functional neuroimaging studies are in progress

13. Are the relationships between task performance and clinical feature(s) known?

i. Partially (see point #1)

14. Is the task feasible for administration across sites?

i. Yes

15. Can the task be used as a stand-alone behavioral task?

i. Yes – most published papers using the task have used it as a stand-alone task

16. What work is needed to get this task ready for use in clinical trials?

i. The task is currently being used in clinical trials for schizophrenia. Preparation for other patient groups may be required.

17. Is the task copyrighted?

i. The task is copyrighted (Emory University and Vanderbilt University) and a license is required for non-academic use (i.e., industry). The task is made freely available for academic research.

References

1. Salamone JD, Cousins MS, McCullough LD, Carriero DL, Berkowitz RJ. Nucleus accumbens dopamine release increases during instrumental lever pressing for food but not free food consumption. Pharmacol Biochem Behav 1994; 49(1): 25-31.
2. Salamone JD, Correa M. The mysterious motivational functions of mesolimbic dopamine. Neuron 2012; 76(3): 470-485.
3. Wardle MC, Treadway MT, Mayo LM, Zald DH, de Wit H. Amping Up Effort: Effects of d-Amphetamine on Human Effort-Based Decision-Making. J Neurosci 2011; 31(46): 16597-16602.
4. Treadway MT, Buckholtz JW, Cowan RL, Woodward ND, Li R, Ansari MS et al. Dopaminergic mechanisms of individual differences in human effort-based decision-making. J Neurosci 2012; 32(18): 6170-6176.
5. Treadway MT, Buckholtz JW, Schwartzman AN, Lambert WE, Zald DH. Worth the 'EEfRT'? The effort expenditure for rewards task as an objective measure of motivation and anhedonia. PLoS ONE 2009; 4(8): e6598.
6. Geaney JT, Treadway MT, Smillie LD. Trait Anticipatory Pleasure Predicts Effort Expenditure for Reward. PLoS ONE 2015; 10(6): e0131357.
7. Yang X-h, Huang J, Zhu C-y, Wang Y-f, Cheung EFC, Chan RCK et al. Motivational deficits in effort-based decision making in individuals with subsyndromal depression, first-episode and remitted depression patients. Psychiatry Res 2014; 220(3): 874-882.
8. Treadway MT, Bossaller NA, Shelton RC, Zald DH. Effort-based decision-making in major depressive disorder: a translational model of motivational anhedonia. J Abnorm Psychol 2012; 121(3): 553.
9. Barch DM, Treadway MT, Schoen N. Effort, anhedonia, and function in schizophrenia: Reduced effort allocation predicts amotivation and functional impairment. J Abnorm Psychol 2014; 123(2): 387.
10. Fervaha G, Graff-Guerrero A, Zakzanis KK, Foussias G, Agid O, Remington G. Incentive motivation deficits in schizophrenia reflect effort computation impairments during cost-benefit decision-making. J Psychiatr Res 2013; 47(11): 1590-1596.
11. McCarthy JM, Treadway MT, Bennett ME, Blanchard JJ. Inefficient effort allocation and negative symptoms in individuals with schizophrenia. Schizophr Res 2016.
12. Reddy LF, Horan WP, Barch DM, Buchanan RW, Dunayevich E, Gold JM et al. Effort-based decision-making paradigms for clinical trials in schizophrenia: part 1—psychometric characteristics of 5 paradigms. Schizophr Bull 2015: sbv089.
13. Treadway MT, Peterman JS, Zald DH, Park S. Impaired effort allocation in patients with schizophrenia. Schizophr Res 2015; 161(2): 382-385.
14. Wardle MC, Treadway MT, de Wit H. Caffeine increases psychomotor performance on the effort expenditure for rewards task. Pharmacol Biochem Behav 2012; 102(4): 526-531.
15. Floresco SB, St Onge JR, Ghods-Sharifi S, Winstanley CA. Cortico-limbic-striatal circuits subserving different forms of cost-benefit decision making. Cogn Affect Behav Neurosci 2008; 8(4): 375-389.
16. Walton ME, Bannerman DM, Rushworth MF. The role of rat medial frontal cortex in effort-based decision making. J Neurosci 2002; 22(24): 10996-11003.

17. Yang X-h, Huang J, Lan Y, Zhu C-y, Liu X-q, Wang Y-f et al. Diminished caudate and superior temporal gyrus responses to effort-based decision making in patients with first-episode major depressive disorder. Prog Neuropsychopharmacol Biol Psychiatry 2016; 64: 52-59.
18. Hughes DM, Yates MJ, Morton EE, Smillie LD. Asymmetric frontal cortical activity predicts effort expenditure for reward. Soc Cogn Affect Neurosci 2015; 10(7): 1015-1019.

Cognitive Systems Final Report

Cameron Carter, M.D. (chair), Neal Cohen, Ph.D., Jordan DeVylder, Ph.D., Dwight Dickinson, Ph.D., Damien Fair, Ph.D., Marta Kutas, Ph.D., Sohee Park, Ph.D., and Lucina Uddin, Ph.D.

I. General Comments

- In keeping with the cognitive neuroscience emphasis that is at the foundation of RDoC, in our selection of tasks and paradigms we have emphasized construct validity as the most essential of the selection criteria. In the interests of time and because they lack sub-construct specificity, we did not focus on self-report instruments for the measurement of cognitive systems. In addition, where they were relevant we were guided by experiences and findings from the CNTRICS/CNTRACS measurement development initiative.

- We have not attempted to orthogonalize construct or sub-constructs. Thus, certain tasks and paradigms may be appropriate for measuring multiple cognitive sub-constructs.

- As is the case for other RDoC systems a number of the cognitive tasks/paradigms that seem most appropriate for certain cognitive sub-constructs are well-established in the literature and construct-valid, but have not gone through rigorous psychometric testing and are not standardized in a CNTRICS/CNTRACS sense. As such these measures were uniformly rated as high on construct validity but low on the presence of psychometric data or standardized methods of administration such as stimulus presentation times, inter-trial intervals, numbers of trials per condition etc. As such each paradigm that was considered is better described as 'classes' of tasks/paradigms than as specific measures. In some cases there is one widely used exemplar of the class – e.g., the Posner version of spatial and non-spatial cuing paradigms, and we note this in the relevant section. For other newer approaches (such as relational inference, described under Declarative Memory) there are relatively few exemplars. In either case, the version of the task described in the references for recommended paradigms can be considered a good starting point for further developmental work devoted to task optimization and psychometric characterization.

- It is likely to be necessary to parameterize tasks differently for different populations. This can be challenging. For many paradigms, it will be necessary to parameterize the tasks in a way that preserves construct validity across populations, e.g., children and adults.

- One advantage of cognitive neuroscience paradigms is that many have been used in research around the world, and many use non-verbal stimuli, reducing or eliminating linguistic and cross-cultural adaptation issues.

- The workgroup discussed certain psychotic phenomena, particularly hallucinations and delusions. These phenomena have presented special challenges and the field has not converged on any well-specified model to date though there are interesting theories under development (e.g. aberrant reinforcement learning, defective Bayesian prediction, impaired performance monitoring, altered efference copy etc.). The group felt that while important these have been challenging to model within an RDoC cognitive systems framework, and given the current state of research and we opted to not spend time on this. Reinforcement learning models, which are relevant, are being addressed under Positive Valence Systems.

- Among the cognitive systems constructs in the existing matrix, language was less well elaborated than the other constructs. Given the specialized nature of the field of linguistics as well as evolving concepts that invoke interactions between cognitive and linguistic systems we felt that it identifying constructs and paradigms from this domain would be best served by a group with more expertise in the area. A new review is noted that may provide some useful context for this discussion: *Elvevag, Cohen et al., An examination of the language construct in NIMH's research domain criteria: Time for reconceptualization! Am J Med Genet B Neuropsychiatr Genet. 2016 Mar 10*

- Traditional cognitive tasks, such as those often used in neuropsychological batteries, were discussed generally, and specifically for some constructs. For the most part, these were judged to lack cognitive construct specificity and were not recommended for RDoC purposes.

II. Organization of the Domain

- There is considerable variation in the elaboration and disaggregation of cognitive domains and constructs in the existing RDoC Matrix. Some are in need of updating in light of current cognitive neuroscience. In this context we did suggest a further elaboration of the construct of attention. Specifically, we identified three important subdomains; controlled versus automatic attention, capacity limitation and interference control, and sustained attention or vigilance.

- As is the case in cognitive neuroscience theory there is considerable overlap among certain domains within the RDoC cognitive systems domain, e.g., between working memory, attention and cognitive control. Similarly, there is overlap between RDoC cognitive systems and other RDoC domains. For example, vigilance is a sub-construct of attention and also an index of arousal. Reinforcement learning is an important form of learning and memory but is being addressed by positive valence systems. Language was initially grouped as cognitive systems construct, but is also central to social behavior.

- Possible additional RDoC domains/constructs were discussed. These included reasoning/inference and future simulation which are emerging areas in cognitive neuroscience with strong clinical relevance.

III. Recommended Tasks

All of the tasks below are recommended for inclusion. As noted above and for other non-cognitive domains while construct validity is strong for all of these measures some will need substantial development in terms of optimization and psychometric evaluation.

Attention

- The paradigms listed for the attention domain are 'classes' of paradigms, as described above. They share the characteristics of wide-use, well-developed literature, and high construct validity. While attention has been very widely studied in basic cognitive neuroscience for several decades, with one exception (the Attention Networks Task, see below), there are no standardized, psychometrically refined versions of these tasks.

  Overt/Covert or Bottom-up v. Top down:

  - Spatial and non-spatial cuing tasks, including Posner versions of the tasks[1] and the Attention Networks (ANT) Task[2]. Limited psychometric data are available for the latter.
  - Visual search paradigm[3]

  Capacity and Interference Control:

  - Attentional blink during rapid serial visual presentation[4]
  - Dual task paradigms, including versions developed by Pashler[5]

  Vigilance:

  - Various tasks have 'catch' trials built in that permit investigators to separate attention lapses from other effects of interest (e.g., change detection working memory tasks, perceptual threshold effects) – attention lapses index vigilance. [6]
  - There is also a class of tasks that have been developed to evaluate attention lapsing or "mind wandering" during a variety of laboratory and everyday activities. These tasks include probes to index subjects awareness of lapsing which have been shown to correlate with objective measures of task performance as well as neural measures of task related brain activity[7]

Perception (Visual)

- We focused our discussion on visual perception – because neural substrates differ among perceptual systems and because the science of visual perception is more well-developed than the science relating to other perceptual systems
- Key issues for perception tasks and paradigms include control for visual acuity, lapses of attention, and working memory impairment.
- For example, CNTRACS work has shown that removing schizophrenia subjects who fail attention lapse 'catch' trials eliminates group differences on the contrast contrast task (see below). This poses both construct validity and measurement concerns.

- Some of the tasks recommended below have been optimized and psychometrically characterized by CNTRACS (Contrast-contrast task, JOVI) or in the case of contrast sensitivity by the MARS company. The remaining should be considered "classes" of paradigms with the referenced versions serving as good starting points for further development into more standardized and reliable measures.

  Perceptual sensitivity threshold:

  - Contrast sensitivity paradigms, such as the versions developed by Mars (http://www.marsperceptrix.com), is proprietary
  - A class of tasks with and a substantial literature also exists, but which haven't been standardized or psychometrically tested

  Surround suppression:

  - These are tasks indexing the effectiveness of lateral inhibition in the visual system.
  - Contrast-Contrast task[6]
    - Good construct validity
    - Optimized and psychometrically refined for adult subjects through CNTRACS[8]
    - This optimized version which includes catch trials to control for attention lapsing is available through the CNTRACS website (http://cntracs.ucdavis.edu/)
    - Relation to clinical conditions unclear – no group differences in schizophrenia after controlling for attention in CNTRACS multi-site study[6]

  Visual integration: Tasks measure the active integration of visual features into percepts.

  - Jittered orientation visual integration task (JOVI)[9]
    - Good construct validity
    - Optimized and psychometrically refined for adult subjects through CNTRACS[9]
    - This optimized version which includes catch trials is available through the CNTRACS website (http://cntracs.ucdavis.edu)
    - Sensitive to group differences[9]

Declarative Memory

- Although there are many memory systems, declarative memory was identified in the original cognitive systems workgroup meeting as the best memory target for psychopathology research within the RDoC framework. We maintained that focus.

  Relational memory

  - The processes involved in memory for stimuli/events and how they were associated with coincident context, stimuli, or events.
  - Relational and item-specific encoding task (RISE)[10]
    - Good construct validity for both relational and item memory performance
    - Optimized and psychometrically refined for adult subjects through CNTRACS

- This optimized version is available through the CNTRACS website (http://cntracs.ucdavis.edu/)
- Good evidence of impairment in clinical groups (schizophrenia) [10].
- Evidence of relationship to everyday functioning in schizophrenia[11]

Associative inference

- An emerging class of paradigms that has good construct validity for the operation of relational memory and the ability to infer new relationships between learned items based upon their relationships with other items acquired during learning.
- This is a developing literature and an with established canonical paradigm but the measures are not yet standardized and have unknown psychometric characteristics[12]. Hence this paradigm is recommended as a construct valid measure that, like many others in the cognitive domain, will need further development of an optimized version that has been psychometrically characterized

Paired-associates learning

- Various measures available (e.g., from Wechsler Memory Scale)
- Standardized with reasonable psychometrics
- The Wechsler version is an option however it has fewer items and less precision than other tasks/paradigms. Experimental tasks are fairly widely used in the literature and developing a standardized task would be straightforward and is recommended.

Pattern separation:

- The ability to distinguish previously presented items from very similar foils (i.e., more challenging than memory tasks involving a simple 'old v. new' distinction). A newer and more computationally specified aspect of Declarative memory that has the advantage of:
  - More sensitive to aging effects and dementia than old v. new paradigms
  - Known relationships to memory circuitry in brain
- Mnemonic Similarity Test, Yassa and Starke[13]
  - Recommended as construct valid but psychometrics not known and will have to be established, in order to have an optimal standardized task.

Cognitive Control

- This construct includes processes needed to maintain goal directed performance and overcome prepotent and habitual responding
- There is substantial sub-construct and task/paradigm overlap with working memory domain

Response inhibition:

- Paradigms listed for this sub-construct are 'classes' of paradigms, as described in our general comments. They share the characteristics of wide-use, well-developed

literature, and high construct validity. We are not aware of standardized, psychometrically refined versions of these tasks.

- Go/No-go tasks[14]
- Stop signal tasks[15]

Goal maintenance (or preparatory cognitive control):

- AX and DPX continuous performance tests
  - Strong construct validity[16]
  - Optimized and psychometrically refined for adult subjects through CNTRACSs[8]
  - Optimized versions are available through the CNTRACS website (http://cntracs.ucdavis.edu/)
  - Evidence of impairment in clinical conditions[17]
  - Related to functional measures
- Preparing to overcome prepotency task (POP)
  - Good construct validity as a goal maintenance/proactive cognitive control measure[18]
    - Psychometrics have not been characterized
    - A good choice for use with impaired populations (psychotic disorders, autism) and children due to simple task structure.[19,20]

Performance monitoring (or dynamic control): Post error and post-conflict adjustments

- Each of these three classes of tasks/paradigms commonly used to measure this contruct is widely-used with well-developed literature, and reasonable construct validity. Psychometric development and optimization are needed except for flankercla[21]/post conflict adjustments measure. Versions exist for use in children[22]. One example is the NIH Toolbox Flanker task, which has been standardized for all ages and psychometrically tested. There are questions about whether it includes sufficient numbers of trials but it might be possible to compute trial to trial and post error adjustments (3 minutes duration).
  - Flanker task versions
  - Simon task versions
  - Stroop task versions[23]

Working Memory

- Working memory has been extensively studied in the cognitive neuroscience and individual differences literature and the measures that have been used for the most part have not been standardized or psychometrically characterized. The exceptions to this are the AX/DPX measure of active maintenance and the change detection measure for working memory capacity that have been developed by CNTRACS.
- As noted above, there is substantial sub-construct and task/paradigm overlap with cognitive control domain

Capacity:

- Change detection
  - Various versions and well-studied paradigm with good construct validity[24]
  - A standardized, psychometrically tested version is in development and will be available through CNTRACS in roughly one year

Flexible updating:

- NBack[25]
  - Many versions and well-studied paradigm with reasonable construct validity
- Self-ordered pointing[26]
  - Widely-used with well-developed literature, and reasonable construct validity.

Active Maintenance:

- Match to sample tasks
  - Various versions and well-studied paradigm with good construct validity[27]
  - Widely used in animal models from primates to birds
- Sternberg tasks
  - Various versions and well-studied paradigm with good construct validity [28]
- Change detection (see above)[24]
- AX and DPX continuous performance tests (see above)

Interference control:

- Related to active maintenance, but emphasizing the extra demands and effort associated with resisting distraction or lapsing attention
  - NBack (see above)
  - Sternberg tasks (see above) [28]
    - Some versions of these tasks have interference built in.

IV. Tasks that are not recommended

Construct Attention

- Mismatch Negativity was mentioned as a fairly well-developed auditory perception paradigm, but it does not yield a behavioral performance index and it was not considered further for current purposes. Also we opted to focus on visual attention for the reasons specified above.

Construct Relational Memory

- Paired Associated Learning form the CANTAB was discussed but not recommended due to low construct validity for item versus relational learning and because it is proprietary.

Construct Working Memory, Flexible Updating

- Letter number sequencing
  - Well-known paper and pencil task that has been standardized and psychometrically refined (e.g., for the WAIS IQ battery
  - Lacks experimental refinement but may be useful to get an approximate index of this sub-construct though not without controversy e.g. [29]

V. References

1. Carter CS, Robertson LC, Chaderjian MR, Celaya LJ, Nordahl TE. Attentional asymmetry in schizophrenia: controlled and automatic processes. *Biological Psychiatry*. 1992;31(9):909-918. doi:10.1016/0006-3223(92)90117-I.
2. MacLeod JW, Lawrence MA, McConnell MM, Eskes GA, Klein RM, Shore DI. Appraising the ANT: Psychometric and Theoretical Considerations of the Attention Network Test. *Neuropsychology*. 2010;24(5):637-651. doi:10.1037/a0019803.
3. Gold JM, Fuller RL, Robinson BM, Braun EL, Luck SJ. Impaired top–down control of visual search in schizophrenia. *Schizophrenia Research*. 2007;94(1-3):148-155. doi:10.1016/j.schres.2007.04.023.
4. Mathis KI, Wynn JK, Breitmeyer B, Nuechterlein KH, Green MF. The attentional blink in schizophrenia: Isolating the perception/attention interface. *Journal of Psychiatric Research*. 2011;45(10):1346-1351. doi:10.1016/j.jpsychires.2011.04.002.
5. Nuechterlein KH, PASHLER HE, SUBOTNIK KL. Translating basic attentional paradigms to schizophrenia research: Reconsidering the nature of the deficits. *Development and Psychopathology*. 2006;18(03):831-851. doi:10.1017/S095457940606041X.
6. Barch DM, Carter CC, Dakin SC, et al. The Clinical Translation of a Measure of Gain Control: The Contrast-Contrast Effect Task. *Schizophr Bull*. November 2011:sbr154. doi:10.1093/schbul/sbr154.
7. Smallwood J, Schooler JW. The Science of Mind Wandering: Empirically Navigating the Stream of Consciousness. *http://dxdoiorg/101146/annurev-psych-010814-015331*. January 2015. doi:10.1146/annurev-psych-010814-015331.
8. Strauss ME, McLouth CJ, Barch DM, et al. Temporal Stability and Moderating Effects of Age and Sex on CNTRaCS Task Performance. *Schizophr Bull*. 2014;40(4):835-844. doi:10.1093/schbul/sbt089.
9. Silverstein SM, Keane BP, Barch DM, et al. Optimization and Validation of a Visual Integration Test for Schizophrenia Research. *Schizophr Bull*. 2011;38(1):sbr141–134. doi:10.1093/schbul/sbr141.
10. Ragland JD, Ranganath C, Barch DM, et al. Relational and Item-Specific Encoding (RISE): Task Development and Psychometric Characteristics. *Schizophr Bull*. 2012;38(1):114-124. doi:10.1093/schbul/sbr146.
11. Sheffield JM, Gold JM, Strauss ME, et al. Common and specific cognitive deficits in schizophrenia: relationships to function. *Cogn Affect Behav Neurosci*. 2014;14(1):161-174. doi:10.3758/s13415-013-0211-5.
12. Schlichting ML, Zeithamova D, Preston AR. CA1 subfield contributions to memory

integration and inference. *Hippocampus*. 2014;24(10):1248-1260. doi:10.1002/hipo.22310.

13. Bakker A, Kirwan CB, Miller M, Stark CEL. Pattern Separation in the Human Hippocampal CA3 and Dentate Gyrus. *Science*. 2008;319(5870):1640-1642. doi:10.1126/science.1152882.
14. Boucher L, Palmeri TJ, Logan GD, Schall JD. Inhibitory control in mind and brain: An interactive race model of countermanding saccades. *Psychological Review*. 2007;114(2):376. doi:10.1037/0033-295X.114.2.376.
15. Luijten M, Machielsen MWJ, Veltman DJ, Hester R, Haan L de, Franken IHA. Systematic review of ERP and fMRI studies investigating inhibitory control and error processing in people with substance dependence and behavioural addictions. *169*. 2014. doi:10.1503/jpn.130052.
16. Lopez-Garcia P, Lesh TA, Salo T, et al. The neural circuitry supporting goal maintenance during cognitive control: a comparison of expectancy AX-CPT and dot probe expectancy paradigms. *Cogn Affect Behav Neurosci*. 2015;16(1):164-175. doi:10.3758/s13415-015-0384-1.
17. Henderson D, Poppe AB, Barch DM, et al. Optimization of a Goal Maintenance Task for Use in Clinical Applications. *Schizophr Bull*. 2012;38(1):104-113. doi:10.1093/schbul/sbr172.
18. Snitz BE, MacDonald A III, Cohen JD. Lateral and medial hypofrontality in first-episode schizophrenia: functional activity in a medication-naive state and effects of short-term atypical antipsychotic treatment. *American Journal of …*. 2005.
19. Solomon M, Ozonoff SJ, Cummings N, Carter CS. Cognitive control in autism spectrum disorders. *International Journal of Developmental Neuroscience*. 2008;26(2):239-247. doi:10.1016/j.ijdevneu.2007.11.001.
20. Solomon M, Ozonoff SJ, Ursu S, et al. The neural substrates of cognitive control deficits in autism spectrum disorders. *Neuropsychologia*. 2009;47(12):2515-2526. doi:10.1016/j.neuropsychologia.2009.04.019.
21. Clayson PE, Larson MJ. Psychometric properties of conflict monitoring and conflict adaptation indices: Response time and conflict N2 event-related potentials. *Psychophysiology*. 2013;50(12):1209-1219. doi:10.1111/psyp.12138.
22. Mullane JC, Corkum PV, Klein RM, McLaughlin E. Interference Control in Children with and without ADHD: A Systematic Review of Flanker and Simon Task Performance. *Child Neuropsychology*. June 2009. doi:10.1080/09297040802348028.
23. Kerns JG, Cohen JD, MacDonald AW, Cho RY, Stenger VA, Carter CS. Anterior Cingulate Conflict Monitoring and Adjustments in Control. *Science*. 2004;303(5660):1023-1026. doi:10.1126/science.1089910.
24. Barch DM, Moore H, Nee DE, Manoach DS, Luck SJ. CNTRICS Imaging Biomarkers Selection: Working Memory. *Schizophr Bull*. November 2011:sbr160. doi:10.1093/schbul/sbr160.
25. Jonides J, Schumacher EH, Smith EE, et al. Verbal Working Memory Load Affects Regional Brain Activation as Measured by PET. *http://dxdoiorg/101162/jocn199794462*. January 2008. doi:10.1162/jocn.1997.9.4.462.
26. Gillett R. Assessment of Working Memory Performance in Self-Ordered Selection Tests.

Cortex. 2007;43(8):1047-1056. doi:10.1016/S0010-9452(08)70702-0.

27. Horwitz B, Tagamets MA. Predicting human functional maps with neural net modeling. *Human brain mapping*. 1999.
28. Nelson JK, Reuter-Lorenz PA, Sylvester C-YC, Jonides J, Smith EE. Dissociable neural mechanisms underlying response-based and familiarity-based conflict in working memory. *PNAS*. 2003;100(19):11171-11175. doi:10.1073/pnas.1334125100.
29. Egeland J. Measuring Working Memory With Digit Span and the Letter-Number Sequencing Subtests From the WAIS-IV: Too Low Manipulation Load and Risk for Underestimating Modality Effects. *Applied Neuropsychology: Adult*. 2015;22(6):445-451. doi:10.1080/23279095.2014.992069.

Systems for Social Processes Final Report

Jed Elison, Ph.D., William Horan, Ph.D., James Morris, Ph.D. Lynn Paul, Ph.D., Kevin Pelphrey, Ph.D. (chair)

I. GENERAL COMMENTS

We selected a set of behavioral performance, self-report and paradigms to assay various social processes described in the RDOC matrix. Our goal was to identify the best, currently available, tasks. We also sought to identify areas where additional research is needed in order to further develop tasks. The majority of tasks, even if considered the best available options, are in need of additional work to support their use across RDoC projects. In particular, all of the tasks we identified need additional refinement (e.g., psychometric properties, norms) for use across pediatric, adolescent, and adult healthy/clinical populations, as well as the development of normative growth curves for the typical development of targeted neural systems. On virtually every paradigm, task, self-report we recommend more research to establish psychometric properties, norms, growth charts of longitudinal developmental changes, development of age-appropriate paradigms and application to clinical groups.

II. ORGANIZATION OF THE DOMAIN

We recommend additional work specifically on the following issues/constructs/tasks:

- Affiliation Construct – The field needs development of methods to assess reciprocal dyadic interaction of established dyads (parent-child; spouses) and dyadic interaction of an individual and stranger.
- Consider Rejection Sensitivity and Social Motivation as key subconstructs under Attachment and Affiliation in the RDOC matrix.
- The construct of Social Communication Initiation (Faces) needs development of both methods of eliciting emotions and methods for measuring facial expression (the latter is ripe for development of new technology).
- For the broader construct of Social Communication (i.e. not face-specific), there is a significant need for development of techniques / instruments that capture the dimensionality of functioning across the life span and populations and instruments that optimize ecological validity.
- The construct of Understanding Mental States needs work focused on taxonomy and task development particularly at higher levels of complexity (inference, irony).
- Affective/Social Touch – Studies are needed to further develop our understanding of this important aspect of social cognition.
- The over-representation of psychodynamically-inspired attachment paradigms (e.g. strange-situation) was concerning because of the general lack of support for the foundational tenets of psychodynamic theory. This is tasks and paradigms emerging from the theory challenging, as they tend to link directly too, and reify the theoretical claims.

III. RECOMMENDED TASKS

Assessment of Measures: We evaluated a series of paradigms under each of the constructs, and rated a selection of the most promising of these tasks. Our ratings can be found below in Appendix SSP-I. A question mark indicates a lack of knowledge / data for the criteria, and thus an opportunity for additional research. Rating the task implies that we are recommending this task as the current best option (even though some are still quite poor options).

A. AFFILIATION AND ATTACHMENT: NONE

Rejection Sensitivity (new proposed subconstruct under Affiliation and Attachment):

Cyberball is the state-of-the-art paradigm for the measurement of Rejection Sensitivity. But it needs development in order to make it useful as a behavioral measure.

- Recent meta-analysis[1] provide extensive information about this task.
- This task has been used in a repeated configuration (essential for longitudinal studies) and findings indicated good repeatability[2] but this may differ by subject cohort.
  - There needs to be more work on test-retest reliability across subject groups and studies of development in children and adolescents.
  - The most consistent outcomes in Cyberball are measured via fMRI, with limited dimensionality of behavioral outcome
  - Elements of the task can be standardized but thus far, the tasks are quite variable.
  - We need work comparing behavioral outcome paradigms and establishing a standardized behavioral implementation of Cyberball.

Social Motivation – Approach / Avoidance – (new proposed subconstructs under Attachment and Affiliation):

We recommend the ***One-Arm Bandit Task***[3] as a measure of social motivation – i.e., to approach social reward or avoid social punishment/threat. This paradigm measures learning in response social feedback (happy vs. angry face stimuli).

- This task hasn't been used widely yet so needs more research to establish psychometric properties, norms, development of age-appropriate paradigms and application to clinical groups.

We recommend the ***Multidimensional Scale of Perceived Social Support*** (MSPS)[4] Questionnaire

- The MSPS has been used with various populations (4 ethnic groups) and the reliability of measure has been reported down to age 3[5]. It has good dimensionality and measures measures state, but needs work on test-retest reliability.

B. SOCIAL COMMUNICATION

Reception of Facial Communication

- Facial emotion (static faces): The **ER40 - Penn Emotion Recognition Test**[6] is recommended.
- Joint attention: The **Gaze Cuing** task[7] is the current gold standard. The task is well established but not well standardized. The effect is clear, but the task needs more work on standardization and the development of norms.

Production of Facial Communication: **NONE** are ready for "prime-time".

Non-Facial Communication (Merged Perception and Initiation)

- We recommend changing the RDOC matrix to merge perception and initiation.
- This area is ripe for development using technology.
- We recommend the **Social Responsiveness Scale**[8] **as a questionnaire**. This is a psychometrically robust measure for various ages. It has good dimensionality across the whole population. It measures a construct that is not static (e.g. weight). It is not a behavioral performance measure, however. Instead it is a report completed by a close other (e.g., spouse or parent).
- We recommend the **TASIT 1**[9] for measuring the perception of emotions presented through multiple modalities**:**.
- There was extensive discussion about the subcategories and difficulty grouping topics: vocal paralinguistic (e.g. pace, prosody, pitch, volume); interpersonal features (e.g. turn-taking, proxemics). This area needs extensive development; specifically related to instruments that capture dimensionality of functioning across age range and populations and instruments that optimize ecological validity.

C. PERCEPTION AND UNDERSTANDING OF SELF

Self-Knowledge

- We recommend the **Self-Referential Memory**[10] task**.** Tasks of this kind have been used mostly in fMRI & ERP paradigms. However, they suffer as behavioral tasks because of the absence of clear behavioral data from these paradigms.

D. PERCEPTION AND UNDERSTANDING OF OTHERS

Animacy Perception

- The current gold-standard for the measurement of Animacy Perception is **Point-Light Displays of Biological Motion**[11]**.** This paradigm needs work on standardization of tasks, and creation of normative data, including growth curves of development of brain mechanisms for the perception of biological motion.
- We also noted the use of **Animations** (e.g. Castelli goal-directed versions[12]) as promising, but in need of standardization.

Action Perception

- We discussed the **How of Why/How Task**[13] (http://www.bobspunt.com/whyhowlocalizer) , noting it need development across many domains, but recommending this task as the best available currently.

- Simple imitation tasks have been employed, but these are normally used as a brain measure with ceiling of behavioral performance.
- Similarly, for Action Perception, especially in terms of actions towards goals, we know of extensive neuroimaging work, but in the absence of a clear behavioral task.

Understanding Mental States

- Logical/Physical Perspective Taking (e.g. False Belief, **Hinting Task,** Stories from Everyday Life). We recommend the **Hinting Task**[14] as a current best option.
- Mental/Emotional Perspective Taking (e.g. Empathic Accuracy, False Belief, TASIT 2 & 3, **Reading Mind in the Eyes**[15]). We recommend the reading **Reading the Mind in the Eyes Task** as a current best option.

IV. TASKS THAT REQUIRE FURTHER EVALUATION

Self-Knowledge

- We discussed Self-Relevant processing (e.g. **Self-Referential Memory**[10]). We noted these tasks have been used mostly in fMRI & ERP paradigms. As such, they suffer as behavioral tasks because of the absence of clear behavioral data from these paradigms.
- We discussed Identification of own emotional states (e.g. TAS-20).
- We considered Reality Testing (e.g. source memory paradigms).
- We discussed the comparison of Self and Other ratings of socially-relevant functions (BRIEF, ABAS).

Production of Facial Communication: NONE

- Spontaneous facial emotion generation: There is not a standardized paradigm for eliciting facial emotion (for examples, see[16,17]). There are various methods for measurement such as FACS (http://www.paulekman.com/facs/ or FACES (socrates.berkeley.edu/~akring/FACES%20manual.pdf), plus EMG. We need development of standardized methods for eliciting emotion and simplified systems for measuring facial expression.
- Mimicry / imitation of emotional expression: We need development of standardized methods for eliciting imitation/mimicry and simplified systems for measuring facial expression.
- Joint attention: There are several excellent researchers who have emphasized this area in children including Peter Mundy and Michael Tomasello. This is a very important construct that is has been developed for ages 9-24 months, but it needs development and measurement the for entire relevant age range, including modification of the construct to include, perhaps, joint intentions, intention sharing in older children and even adults.
- We noted that the area of meta-cognition of performance needs development.

V. TASKS THAT ARE NOT RECOMMENDED

We reviewed the tasks currently suggested in the RDOC matrix. From this review, we recommend *eliminating* from the matrix the: 1) separation-reunion (e.g., Strange Situation); 2) Still Face; and 3) Ford Corollary Discharge paradigms. Separation-reunion paradigms are limited by an inherent reliance on a particular set of theoretical assumptions (i.e., attachment-theory) that are widely debated. The Still Face paradigm is very specific to one developmental epoch and has little predictive utility. The Ford paradigm does not utilize a behavioral output, and is thus of limited use. This recommendation is not meant to imply that we endorse all the rest of the tasks listed in the matrix in the social domain—only that we identified particularly significant problems with the ones we recommend here for removal.

We discussed but excluded from further consideration the Social Network Index[18]. On the one hand, this isn't a behavioral performance measure, making it less well suited for the RDOC approach. However, this measure is associated with amygdala volume[19]. We discussed the use of this measure to assess affiliation, but decided not to recommend this task because it is an index of social outcomes that are informed by multiple processes.

References

1. Hartgerink CHJ, van Beest I, Wicherts JM, Williams KD. The Ordinal Effects of Ostracism: A Meta-Analysis of 120 Cyberball Studies. Van Yperen NW, ed. *PLOS ONE*. 2015;10(5):e0127002. doi:10.1371/journal.pone.0127002.
2. Bolling DZ, Pitskel NB, Deen B, et al. Dissociable brain mechanisms for processing social exclusion and rule violation. *NeuroImage*. 2011;54(3):2462-2471. doi:10.1016/j.neuroimage.2010.10.049.
3. Lin A, Adolphs R, Rangel A. Social and monetary reward learning engage overlapping neural substrates. *Soc Cogn Affect Neurosci*. 2012;7(3):274-281. doi:10.1093/scan/nsr006.
4. Zimet GD, Dahlem NW, Zimet SG, Farley GK. The multidimensional scale of perceived social support. *Journal of personality assessment*. 1988;52(1):30-41.
5. Bruwer B, Emsley R, Kidd M, Lochner C, Seedat S. Psychometric properties of the Multidimensional Scale of Perceived Social Support in youth. *Comprehensive Psychiatry*. 2008;49(2):195-201. doi:10.1016/j.comppsych.2007.09.002.
6. Erwin RJ, Gur RC, Gur RE, Skolnick B, Mawhinney-Hee M, Smailis J. Facial emotion discrimination: I. Task construction and behavioral findings in normal subjects. *Psychiatry Research*. 1992;42(3):231-240. doi:10.1016/0165-1781(92)90115-J.
7. Friesen CK, Kingstone A. The eyes have it! Reflexive orienting is triggered by nonpredictive gaze. *Psychonomic Bulletin & Review*. 1998;5(3):490-495. doi:10.3758/BF03208827.
8. Constantino JN, Davis SA, Todd RD, et al. Validation of a Brief Quantitative Measure of Autistic Traits: Comparison of the Social Responsiveness Scale with the Autism Diagnostic Interview-Revised. *J Autism Dev Disord*. 2003;33(4):427-433. doi:10.1023/A:1025014929212.
9. McDonald S, Flanagan S, Rollins J, Kinch J. TASIT: A New Clinical Tool for Assessing Social Perception After Traumatic Brain Injury. *The Journal of Head Trauma Rehabilitation*.

2003;18(3):219.
10. Kelley WM, Macrae CN, Wyland CL, Caglar S, Inati S, Heatherton TF. Finding the Self? An Event-Related fMRI Study. *http://dxdoiorg/101162/08989290260138672*. March 2006. doi:10.1162/08989290260138672.
11. Björnsdotter M, Wang N, Pelphrey K, Kaiser MD. Evaluation of Quantified Social Perception Circuit Activity as a Neurobiological Marker of Autism Spectrum Disorder. *JAMA Psychiatry*. 2016;73(6):614-621. doi:10.1001/jamapsychiatry.2016.0219.
12. Castelli F, Frith C, Happé F, Frith U. Autism, Asperger syndrome and brain mechanisms for the attribution of mental states to animated shapes. *Brain*. 2002;125(8):1839-1849. doi:10.1093/brain/awf189.
13. Spunt RP, Adolphs R. Validating the Why/How contrast for functional MRI studies of Theory of Mind. *NeuroImage*. 2014;99:301-311. doi:10.1016/j.neuroimage.2014.05.023.
14. Corcoran R, Frith CD. Autobiographical memory and theory of mind: Evidence of a relationship in schizophrenia. *Psychological Medicine*. 2003;33(5):897-905. doi:10.1017/S0033291703007529.
15. Vellante M, Baron-Cohen S, Melis M, et al. The "Reading the Mind in the Eyes" test: Systematic review of psychometric properties and a validation study in Italy. *Cognitive Neuropsychiatry*. July 2013. doi:10.1080/13546805.2012.721728.
16. Gross JJ, Levenson RW. Emotion elicitation using films. *Cognition & Emotion*. January 2008. doi:10.1080/02699939508408966.
17. Lang PJ, Bradley MM, Cuthbert BN. International affective picture system (IAPS): Affective ratings of pictures and instruction manual. *Technical report A-8*. 2008.
18. Cohen S, Doyle WJ, Skoner DP, Rabin BS, Gwaltney JM. Social Ties and Susceptibility to the Common Cold-Reply. *JAMA*. 1997;278(15):1232-1232. doi:10.1001/jama.1997.03550150036020.
19. Bickart KC, Wright CI, Dautoff RJ, Dickerson BC, Barrett LF. Amygdala Volume and Social Network Size in Humans. *Nature neuroscience*. 2011;14(2):163. doi:10.1038/nn.2724.

Appendix SSP-I: Ratings for recommended tasks on task criteria

Construct	Task	Valid Measure of Construct	Test-Retest Reliability	Floor/Ceiling / dimensionality	Practice Effects	Longitudinal Stability	Sensitive to w/in person	Standardized Parameters	Can it be used across many sites?	Can it be used with kids or special	Are normative data available?	Are relations to clinical features	Not copyrighted?
	Criteria (Rate each on a scale of 1-5, with 1 = does not do a good job of meeting the criterion;												
	5 = does an excellent job meeting the criterion												
Attachment and Affiliation	1. Direct brain and behavior measures of rejection sensitivity (**Cyberball**, Chat Room)	5	?	2	3	?	?	2	5	5	3	3	5
	2. Social Motivation – Approach / Avoidance (e.g., Social Incentive Delay Task; Dot Probe / Attention Bias; **One-Armed-Bandit Task** and other neuroeconomics tasks)	4	?	4	4	?	?	5	5	3	?	?	5
	3. Self report of need for affiliation / rejection sensitivity (**Multidimensional Scale of Perceived**	5	?	5	NA	?	?	5	5	4	4	?	5

	Social Support; Anhedonia Scale; Rejection Sensitivity Scale; Affiliation Tendency Scale)												
Social Communication													
Reception of Facial Communication	Joint attention, responding (**gaze cuing**);	5	?	5	5	?	?	1	5	5	?	3	5
	Facial Emotion (face scan paths; **ER40**)	5	4	4	5	3	?	5	5	5	4	4	5
Production of Facial Communication	**Joint attention, initiating**	?	?	?	?	?	?	?	?	4	?	?	5
		?	?	?	?	?	?	?	?	?	?	?	?
	Initiate emotional expression (EMG / FACS / FACES with provocative context – IAPS / Affective Computing)												
	Mimicry / Imitate emotional expression (FACS coding / EMG	?	?	?	?	?	?	?	?	?	?	?	5
Reception of Non-Facial Communication	**Social Responsiveness Scale (SRS)**	3	5	5	NA	5	4	5 \	5	5	5	5	2

	Multimodal Social Pragmatics (e.g., vocal: pace, prosody, pitch,; turn-taking, distance, touch, gestures; BLERT; TASIT 1; CASL-PL)	?	?	?	?	?	?	?	?	?	?	?	?
Perception and Understanding of Self													
Agency	Illusions of will (rubber hand)	?	?	?	?	?	?	?	?	?	?	?	?
	Joystick Manipulation (decoupling motor and sensory feedback)												
	Measure of sense of control												
Self-Knowledge	**Self-Referential Memory Paradigm**	4	3	4	4	3	?	4	5	5	?	3	5
	Self-referential effect												
	BRIEF, ABAS – self vs. other rating of knowledge of self												
	Identify own emotional states												
	Reality monitoring												

Perception and Understanding of Others													
Animacy perception	**Point light displays of biological motion**	5	3	4	3	3	4	3	5	5	3	5	5
	Heider and Simmel type films (goal-directed)												
Action Perception	**How part of the How / Why Task (Bob Spunt)**	5	?	?	?	?	?	5	5	5	?	?	5
	Imitation (finger tapping)												
	Contagion (yawning)												
	Action observation – goal directed actions												
Understanding Mental States	Logical / Physical Perspective taking (e.g., **Hinting Task**, False Belief, Stories from Everyday Life)	4	4	4	4	4	3	5	5	3	?	4	5
	Mental/Emotional Perspective Taking (e.g., **Reading the Mind in the Eyes**, False Belief, Empathic Accuracy Paradigm, TASIT parts 2 & 3)	3	2	4	4	5	3		5	4	3	5	5

Arousal and Regulatory Systems Final Report

Theodore P. Beauchaine, Ruth M. Benca, David Brent, Andrew Krystal, Dara Manoach (chair), Uma Vaidyanathan

This report summarizes the meeting of the Arousal and Regulatory Systems (ARS) subgroup of the National Advisory Mental Health Council Workgroup, which was established by the National Institute of Mental Health (NIMH) to develop "a list of recommended tasks for each RDoC construct included in the RDoC matrix". The ARS subgroup was charged with developing a list of tasks for the three constructs in the Arousal and Regulatory Systems domain: (1) arousal, (2) sleep-wakefulnesss and (3) circadian rhythms. Arousal/regulatory systems are defined in the RDoC matrix as, "…**responsible for generating activation of neural systems as appropriate for various contexts, and providing appropriate homeostatic regulation of such systems as energy balance and sleep** (http://www.nimh.nih.gov/research-priorities/rdoc/constructs/arousal-and-regulatory-systems.shtml, June 20th, 2016). Since effortful/active forms of behavior regulation and emotion regulation are not subsumed within this definition, they were not considered. Thus, the ARS subgroup worked directly from constructs already listed in the RDoC matrix (http://www.nimh.nih.gov/research-priorities/rdoc/constructs/rdoc-matrix.shtml), without revising or editing those constructs.

CONSTRUCT: AROUSAL

GENERAL ISSUES: The task of recommending paradigms to assess arousal presented the ARS subgroup with a number of challenges. Arousal is not well-defined in the RDoC matrix, perhaps because the term has historical roots that cut across constructs including orienting, vigilance, attention, motivation, trait and state anxiety, stress responding, and coping, among others. These constructs are subserved by a wide array of neurobiological processes and functions, some cortical, some subcortical, and some peripheral. These neurobiological processes and functions are facilitated/regulated by multiple monoamine and neuroendocrine systems that are distributed across diverse brain regions. Even single monoamine neurotransmitters (e.g., dopamine) can be subdivided into separate systems (e.g., mesocortical, tuberoinfundibular) that are implicated in distinct arousal functions (attention, sleep). Given these complexities, and the tendency of what we think of as distinct psychological functions to be widely distributed across neural networks, it may not be possible to define arousal as "…distinct from motivation and valence…", as outlined in the RDoC matrix (http://www.nimh.nih.gov/research-priorities/rdoc/constructs/arousal.shtml).

This above paragraph illustrates why many contemporary psychophysiologists avoid the term arousal altogether. That said, there is considerable psychophysiological research on constructs, such as those listed above, that can be construed as related to arousal. Historically, much of this work was conducted at the autonomic nervous system (ANS) level. More recently, electroencephalography (EEG) and functional neuroimaging (e.g., fMRI) have been used. . Valid self-report measures are lacking, but would be valuable to develop.

MEASUREMENT ISSUES: Most of the measures considered did not have agreed upon standards for administration or analysis and all would benefit from normative data. All are scalable to varying degrees. All tasks can be affected by multiple factors, many of which cross into other RDoC domains (negative and positive valence, cognition). All require tightly controlled experimental conditions for reliable and externally valid assessment.

AUTONOMIC MEASURES OF AROUSAL

Most autonomic measures, including heart rate (HR), blood pressure, and pupil diameter, are affected by both the sympathetic and parasympathetic nervous systems (SNS, PNS)[1]. This is problematic when attempting to parse relative contributions of bottom-up, largely subcortical, *emotion generation* systems from top-down, largely cortical, *emotion regulation* systems (e.g., prefrontal)[2]. Excitatory sympathetic efference is a better index of arousal, whereas inhibitory parasympathetic efference is a better index of regulation[3]. HR cannot be used to capture either construct because it is represented in two-dimensional space, with SNS activation (low to high) on one axis and PNS activation (low to high) on the other axis[4]. Thus, an almost unlimited number of SNS and PNS combinations can result in a single HR value. For these reasons and others (both functional and anatomical), cardiac pre-ejection period (PEP; assessed using impedance cardiography), a noninvasive index of SNS activation (given appropriate stimulus conditions), is a better index of arousal. PEP shortening, which reflects increased SNS activation, is observed reliably during stress- and emotion-induction tasks, including public speaking and other paradigms in which one's performance is evaluated by outside observers[5,6].

RECOMMENDED MEASURE 1: HEART RATE VARIABILITY

Captures both arousal and regulation. Under appropriate stimulus conditions, high frequency heart rate variability (HRV, >.15 Hz among adults) assesses PNS efference to the heart[7]. Although reductions in HRV are observed in disorders where arousal is implicated in symptom expression (e.g., panic, phobias, and other anxiety disorders; non-suicidal self-injury; hostility), such reductions in HRV are in no way specific to 'disorders of arousal'. In fact, low resting HRV and/or excessive reductions in HRV during emotion evocation are also observed among those with attention problems, autism, callous unemotional traits, conduct disorder, executive function difficulties, psychopathy, and schizophrenia[2]. Strong arguments have been made that HRV is a peripheral index of emotion regulatory processes that are affected by prefrontal mechanisms. Thus, SNS measures are likely better indices of arousal (particularly during social evaluation), whereas HRV is likely a better index of regulation (particularly during emotion evocation).

- **Reliability:** Several studies have evaluated reliability of HRV measures[8-12]. In general, reliability is good for both resting state measures and task measures, when such tasks are tightly controlled (e.g., well trained administrator, no movement, the same task used across assessment points), and when spectral analysis is used. Clear developmental increases in HRV are observed, which are obscured when age-appropriate respiratory frequencies are not used when spectral analyzing R-R time series. Perhaps unsurprisingly, reliability is poor when different tasks are used to evoke HRV reactivity at different time points.
- **Norms:** Although well-established norms exist for resting HRV, no such norms exist for HRV reactivity, largely because there are no established reactivity tasks[13,14]. This is problematic given that arousal is largely a construct of reactivity. However, reactivity depends on a number of factors that are difficult to standardize, including age and other components of the RDoC matrix (see above), age, physical fitness, individual differences in executive function, and cardiovascular health, to name but a few.
- **Other populations:** There has been some work done on populations including children, adolescents, older adults[1,9,12,15-18].

- **Genetics, heritability and molecules** have also been explored[19-22].
- **Brain circuitry correlates:** As recently reviewed by Beauchaine and Thayer[2], extensive neuroimaging studies using PET, SPECT, and MRI indicate that HRV falls under control of an interconnected cortical and subcortical network, including efferent pathways from the medial prefrontal cortex to the PNS. Through this network, the prefrontal cortex provides top-down inhibitory control over subcortical and brainstem systems that regulate autonomic function[23-25]. Altered function of this PFC network is observed across a wide range of psychopathologies[26,27].
- **Use in clinical trials** (yoga RCTs, open CBT trials, emotional regulation training)[28-32]. Clinical trials conducted to date among children, adolescents, and adults indicate that both resting state HRV and HRV reactivity improve in response to clinical interventions for several forms of psychopathology. These changes correlate with improved emotion regulation capabilities. A number of additional clinical trials are underway.
- **Clinical correlates** (depression, bipolar, schizophrenia)[17,33,34]. Clinical correlates of low resting state HRV and excessive HRV reactivity (parasympathetic withdrawal) are wide-ranging and correspond with difficulties with emotion regulation. All of the following psychiatric disorders/clinical syndromes are characterized by low resting state HRV and/or excessive HRV reactivity (for a review see Beauchaine & Thayer, 2015): anxiety, attention problems, autism, callousness, conduct disorder, depression, non-suicidal self-injury, panic disorder, phobias, trait hostility, psychopathy, schizophrenia.

RECOMMENDED MEASURE 2: Electrodermal Responding

Description: Electrodermal responding (EDR), also known as galvanic skin response or skin conductance, is a good peripheral biomarker of low arousal states/traits, but not as good at demarcating high arousal states/traits. It has been used in studies of emotion and cognition since the early 1900s and is thought to index perceived stimulus significance[35]. It is generally measured by indexing changes in conductance that occur in sweat glands in the skin after passing a weak electrical current through electrodes placed on the fingertips. It does not differentiate between affective states, but is more dependent on the arousal value of stimulus. It has both tonic and phasic components. Phasic components are measured as the change in conductance upon presentation of a stimulus, while tonic responses include skin conductance level and skin conductance level (see Boucsein, 2012[36] for a comprehensive review). Some research has shown that it covaries with amygdala activation to external stimuli[37].

Measurement issues: EDR shows high reliability (see [38] for a review) and biometric heritability[39]. However, as with many biomarkers, links to candidate genes have proved tenuous with initial findings not holding up in larger samples.[39] EDR is easy to measure and wearable devices such as Empatica are available but are only now starting to be incorporated into larger lab-based studies. As with other psychophysiological measures, stimulus conditions are important when evaluating whether EDR marks arousal. EDR covaries with amygdala activation to external stimuli in well-controlled experiments. Non-specific fluctuations in the EDR signal may be more sensitive than other indices (e.g., amplitude, rise time, recovery time, etc.). Information on recording standards and response ranges is available[35,40].

Clinical relevance: Decreased EDR has been associated with externalizing disorders such as antisocial behavior and alcohol abuse/dependence concurrently and predictively[41-47]. Studies of schizophrenia are divided with about half showing no response to stimuli during habituation paradigms[35,48]. PTSD is associated with increased electrodermal responding[49,50].

RECOMMENDED MEASURE 3: Pupillometry

Pupillometry refers to the measurement of pupil diameter, including constriction and dilation. This is accomplished using video recording and eye-tracking. Pupil dilation correlates with activation of noradrenergic fibers originating in the locus coeruleus, which supports use of pupillometry as an index of sympathetic nervous system (SNS) function and arousal[51,52]. However, pupil diameter is also determined by PNS innervation. It therefore suffers from the same interpretability problems as HR (see above). Furthermore, pupil dilation is associated with aspects of emotion and cognition, and varies as a function of emotional saliency[53], task difficulty[54], and attention deployment[55]. Thus, changes in pupil diameter are not specific to arousal. Under some circumstances, pupil dilation may more accurately track temporal aspects of attention than imaging methods such as EEG and fMRI[56], [55]. Although pupil size changes reflexively depending on ambient light conditions, pupillary responses are also evoked by mechanisms unrelated to visual perception, which may or may not be recognized consciously[57].

- **-Reliability:** A limited number of studies have assessed reliability of pupillometry. However, existing studies suggest good to excellent test-retest reliability for changes in pupil diameter, depending on the task used and the population being studied[58].
- **-Use in clinical trials:** Data on changes in pupillary responding following treatment for psychopathology are sparse. However, remission following cognitive therapy for depression is associated with low sustained pupillary responses to negative words[59].
- -**Clinical correlates:** Pupillometry has been used most extensively among children, adolescents, and adults with autism spectrum disorder. While viewing others' faces, those with autism show pupillary constriction, in contrast to typical dilatory responses exhibited by mental-age and chronological-age matched controls[60]. Martineau et al.[61] were able to differentiate children with autism from mental- and chronological-age matched controls with 72% accuracy based on reductions in pupil size when viewing neutral faces, virtual faces, and objects. Pupil dilation may mark more general tendencies toward anxiety and depression, at least in certain contexts[62]. Reduced pupil dilation to negative words is associated with depression severity and negative affectivity, and with low levels of positive affectivity[63].

RECOMMENDED MEASURE 4: CARDIAC PRE-EJECTION PERIOD

Although not discussed at the meeting because a committee member with expertise was not in attendance, cardiac pre-ejection period (PEP) should be included in any discussion of putative autonomic indices of arousal. PEP is defined by the time elapsed (ms) between (1) onset of left ventricular depolarization and (2) ejection of blood into the aorta[64] (see Sherwood et al., 1990). Pharmacology blockade studies indicate that PEP changes in response to internal and external stimuli are mediated fully by beta-adrenergic (SNS) mechanisms[65]. Well controlled experiments

demonstrate PEP shortening during heightened arousal states, including those induced by incentive responding, threat, and psychological stress[66,67]. Unlike HRV however, which changes during these conditions and many others (see above), PEP responding is much more specific[68].

- **-Reliability**: Cardiac PEP demonstrates adequate to excellent internal consistency and test-retest reliability. Cronbach's alphas during difficult tasks with social evaluative components—which elicit considerable arousal—are excellent[69]. Furthermore, stability of PEP responding is observed across intervals as long as a decade[70].
- **-Norms**: As reviewed by Zisner and Beauchaine[1], resting PEP increases monotonically as a function of age through early adulthood, after which age-related changes are negligible[71]. Developmental norms for PEP reactivity are more difficult to establish, since different labs use different stimuli to evoke PEP responses. However, research using reward tasks in particular shows no consistent differences across preschool, middle childhood, adolescence, and early adulthood[68,72-74]. In contrast, Quigley and Stifter[75] reported greater PEP reactivity among young adults than among preschoolers in response to a series of reaction time, emotion evocation, and interview tasks. Thus, consistent with recommendations throughout this report, stimulus conditions need to be considered carefully when interpreting autonomic responding, and in evaluating whether autonomic responses represent changes in arousal (see above).
- **-Genetics, heritability, molecules**: Resting PEP, ambulatory recordings of PEP throughout the day, and PEP reactivity to stress are moderately to highly heritable in adolescence and middle age[76,77]. Few molecular genetics studies specifically of PEP/PEP reactivity have been conducted. However, since PEP change is effected through the SNS via beta-adrenergic mechanisms, candidates include the b1- and b2-adrenergic receptor genes (ADRB1, ADRB2)[78], and other genes that affect SNS-linked cardiovascular reactivity.
- **-Brain circuitry correlates**. Since SNS-mediated increases in cardiac output serve to facilitate behavioral mobilization to multiple arousal states, including incentive responding, threat, and stress, no single neural network is responsible for evoking PEP responses. During incentive tasks in particular, PEP shortening likely originates in dopaminergic reactivity within the striatum, which initiates brainstem responding to mobilize a cardiac response[3,79]. During conditions of threat and stress, PEP responding is likely initiated by other, well characterized neural networks (e.g., SAM).
- **-Use in clinical trials**: To date, PEP has not been evaluated in many clinical trials. However, in a recent RCT of a behavioral intervention for early-onset conduct problems, Beauchaine et al.[28] (2013) found main effects of PEP activity and reactivity on treatment outcomes. Although sample-wide improvements in behavior were observed at post-treatment, those who exhibited lengthened cardiac PEP at rest and reduced PEP reactivity to incentives scored higher on measures of conduct problems and aggression both before and after treatment. This is consistent with a low arousal interpretation of conduct problems and treatment response. Moreover, cross sectional research comparing PEP reactivity to stress among currently depressed versus remitted patients indicates blunted PEP reactivity only in those who are currently depressed[80]. Thus, PEP reactivity may be a state dependent marker of clinical depression.

- **-Clinical correlates**: Compared with controls, males with ADHD, oppositional defiant disorder, conduct disorder, and antisocial personality traits exhibit either diminished PEP reactivity to monetary incentives, or no PEP reactivity at all[3]. Similar findings apply to depressed individuals[80]. PEP non-reactivity may therefore mark attenuated mesolimbic reactivity to reward in both externalizing behavior and depression. Diminished PEP reactivity to incentives also provides prospective prediction of substance use initiation and escalation among middle-schoolers[81].

Promising measures requiring further development

- ERPs during sleep (e.g., oddball paradigm)[82]
- EEG beta/gamma/theta activity during sleep and wake[83]

Considered but not recommended

- Hunger Visual Analogue Scale (insufficient data on properties)
- Heart rate (see above)
- Interoception (not a primary measure of arousal)
- CO2 inhalation (used to trigger acute fear; negative valence measure)
- Startle (more a measure of reactivity to negative stimuli)
- Blood pressure (regulation too complex)
- Auditory steady state response (ASSR) measure of gamma band (Referred to the Cognition Group)
- Trier Social Stress Test, Fear Faces, IAPS pictures (These stimuli evoke arousal but are not arousal measures)
- Cortisol and serum/urine norepinephrine levels (insufficient psychometric validity)

COGNITIVE MEASURE OF AROUSAL

RECOMMENDED MEASURE 4: Psychomotor Vigilance Task (PVT)

The PVT requires participants to press a button as soon as a light appears on a screen at random intervals. RT and the number missed button presses are the dependent measures. The PVT requires sustained attention and is thought to reflect alertness. Poor performance is associated with sleep deficit[84]. Performance is also influenced by motivation and circadian factors. Advantages to the PVT are that it has been well studied, has simple metrics, is brief, free from learning effects and easily scored. The animal version that has been used to detect the effects of sleep deprivation[85]. Performance depends on the basal forebrain[86] and can be disrupted with adenosine infusion producing behavioral deficits resembling sleep deprivation[87].

CONSTRUCT: SLEEP-WAKEFULNESS

Polysomnography (PSG): The first four recommended measures depend on polysomnography, the sleep EEG. General issues with PSG include scalability given the expense and time required for sleep studies. It remains to be established under what conditions a nap or home sleep recordings can substitute for nocturnal sleep in the lab. Scoring and artifact rejection can be laborious, but automatic methods exist and are being developed. Scoring and measurement of sleep architecture and quality are highly standardized. A minimum of two sleep sessions is recommended as the first session is generally considered an adaptation night (or nap) that acclimates the participant to the sleep lab and recording.

RECOMMENDED MEASURES 1: Latency to persistent sleep (LPS), Wake time after sleep onset (WASO), Total Sleep Time (TST)[88-90]

- These are all standard well-established measures.
- recommended two night minimum

RECOMMENDED MEASURE 2: Sleep Spindles (87, 96 Rechtschaffen & Kales 1968; Iber et al., 2007**)**

Description: Sleep spindles are a defining oscillation of Stage 2 non rapid eye movement sleep (N2) seen on PSG as 12-15 Hz oscillations lasting 1 to 2 seconds in a waxing waning envelope. Spindles are also seen in N3, but have different characteristics and functional correlates. The most common metrics are sleep spindle number and density. The morphological characteristics of sleep spindles are also often characterized including peak amplitude, sigma power, duration and frequency.

Measurement issues: There are several publicly available automated methods to detect spindles that have been validated against hand scoring by experts. But experts do not have perfect inter-rater reliability[91] and while internally consistent, different methods give rise to different estimates. Sigma power (12-15 Hz), which is the spindle frequency, is often used as a proxy for spindle activity, but correlates only moderately with hand or automatically detected spindles. Spindles have been divided into slow and fast frequency events, but definitions differ. Some papers have defined fast and slow spindles as covering 13.5-15 Hz and 12-13.5 Hz bands, respectively, while others have identified a lower band of spindles from 9-12 Hz. Fast and slow spindles have different scalp topographies, relations with other NREM sleep oscillations and relations with waking cognition, including sleep-dependent memory consolidation. Spindles change over the lifespan[92]. Normative data is soon to be available.

Clinical relevance: spindle activity is highly heritable and is related to the functioning of genes that confer increased risk for schizophrenia[93,94] and other neurodevelopmental disorders including autism[95]. Spindle generation depends on a well-defined physiology and circuitry involving the thalamic reticular nucleus and thalamocortical circuitry that is implicated in psychopathology. Sleep spindles are relevant to cognition and correlate with sleep dependent memory consolidation, IQ and measures of learning potential in health[96] and psychopathology[97,98]. They can be experimentally manipulated in both humans and animals using pharmacological and neurostimulation techniques to improve sleep-dependent memory consolidation[99,100].

RECOMMENDED MEASURE 3: NON - RAPID EYE MOVEMENT SLEEP (NREM) EEG Slow wave activity: A measure of sleep homeostasis (Rate of Decline in NREM EEG Delta Power across night, NREM EEG Average Delta Power)[88,101-115]

Overview: Sleep homeostasis refers to the increase in propensity to sleep that occurs in proportion to the duration of prior wakefulness. The intensity of the propensity to sleep at any given point is referred to as the degree of "homeostatic sleep drive". There is a compelling body of literature in humans and animals indicating that the dynamics of EEG power in the Delta frequency band (typically 0.5-4 Hz) during NREM sleep reflect the degree of homeostatic sleep drive that has built up at the time of sleep onset and the dissipation of this drive that occurs with sleep. These studies demonstrate that NREM EEG Delta Power in the early part of the night and the rate of decline in NREM EEG Delta power over the night increase proportionally with the duration of prior waking and are decreased by manipulations that decrease homeostatic sleep drive such as extending sleep and napping.

Measurement issues: Studies establishing the relationship between NREM EEG Delta Power dynamics and homeostatic sleep drive have employed a number of standard methods. These include the use of standard methods for computing EEG Delta Power employing the Fast Fourier Transform (FFT) and identifying NREM sleep using standard sleep staging criteria[88,101]. However, studies including NREM EEG Delta Power dynamics have varied in a number of key aspects of methodology, which remain unstandardized. These include the number and location of scalp EEG electrodes utilized in estimating NREM EEG Delta Power, the methods for identifying and removing data contaminated by artifact, and the range of frequencies which define the Delta frequency band.

RECOMMENDED MEASURE 4: Multiple Sleep Latency Test (MSLT): A Measure of Daytime Sleepiness

Overview: The Multiple Sleep Latency Test (MSLT) is a standardized laboratory assessment of the degree of daytime sleepiness. It is measured as the propensity to fall asleep when presented with an opportunity in an environment conducive to sleep. Subjects are given 4-5 opportunities to nap in a quiet, dark room spread across the day. For each nap opportunity, the time to sleep onset is determined from PSG using standard scoring criteria[88]. The average time to sleep onset is the measure of sleep propensity. Because the test requires PSG monitoring in the laboratory the night before testing, subjects must spend a night and a subsequent whole day in the laboratory. Although methods have varied, a standardized protocol for the MSLT has been proposed and widely adopted[116].

Measurement issues: The MSLT has good face validity as a measure of sleepiness and it has been established to have good test/retest reliability. Convergent validity has been established in that it reflects the effects of sleep deprivation and the effects of sedating drugs. Normative data have also been established[117]. Its main limitations in terms of measurement issues are that there are floor and ceiling effects that affect application and that it has significant inter-individual variability such that it is inconsistent in distinguishing healthy controls without sleepiness complaints from individuals with disorders of excessive sleepiness. As such, it is a better measure of within-subject change in sleepiness than an absolute sleepiness measure.

RECOMMENDED MEASURE 5: Insomnia severity index (ISI)

This is the only self-report measure that we recommend. Better validated self-report measures of sleep and sleepiness are needed since widely used measures have poor psychometrics (e.g., Sleep logs/diaries are not standardized or validated). The ISI is limited to insomnia assessment, it is not a general sleep-wake measure[118].

RECOMMENDED MEASURE 6: Finger tapping motor sequence task (MST)

Description: The MST is a measure of sleep-dependent daytime function[119]. It assesses the restorative and transformative properties of sleep on cognition measured during wake. It is the most well-validated measure of sleep-dependent memory consolidation. The MST requires participants to repeatedly type a 5-digit sequence (*e.g.*, 4-1-3-2-4) on a keyboard with the left hand, "as quickly and accurately as possible" for twelve 30 s trials separated by 30 s rest periods. Participants train before sleep and are tested on an additional 12 trials after sleep. The primary outcome measure is *overnight improvement* calculated as the percent increase in correctly typed sequences from the last three training trials to the first three test trials[120]. The MST taps procedural learning and memory. Overnight improvement on the MST correlates with sleep spindle density[121-123] and, in one study, changes in sigma activity in the supplementary motor area as measured by MEG[124]. Patients with schizophrenia and depression generally perform as well as controls in the initial session in terms of the amount and proportion of learning, but show a specific deficit in overnight improvement[97,98,125], that in schizophrenia, correlates with a sleep spindle deficit[126]. Overnight improvement has been linked to prefrontal hippocampal connectivity during learning[127].

Measurement issues: Administration is computerized. Performance may be affected by keyboarding experience. Ideally, the task is administered on consecutive days. Measurement may be highly variable in some participants and investigators have different methods of eliminating outlying responses.

Considered but not selected:

- Word-pair associates learning: A well-validated measure of sleep-dependent declarative memory consolidation. A potential issue in psychopathology research is that participants often have declarative encoding deficits that render it difficult to attribute poor recall to sleep[128].
- Maintenance of Wakefulness Test (MWT) (probes ability to remain awake, but insufficient psychometric data/norms)
- Actigraphy (not a reliable measure of sleep, included under circadian rhythms)

CONSTRUCT: CIRCADIAN RHYTHMS

General concerns: Most measures of circadian rhythms can be influenced by circadian entraining factors such as light, activity, feeding and timing of sleep. For circadian outputs that are also influenced by homeostatic sleep factors, such as cognitive performance, duration of wakefulness can also effect measurement. As a result, circadian rhythms in humans have often been measured under one of several protocols, including time isolation (subjects placed in an environment isolated from time cues, no longer commonly used), forced desynchrony (subjects forced to follow rest-activity schedules that are too short or too long to permit entrainment), and constant routine (subjects remain awake and semi-recumbent in dim light, with nourishment provided at a uniform level throughout)[129-131]. Core body temperature rhythm measurement, which used to be considered the gold standard, is no longer frequently used due to its invasiveness (subjects need to wear rectal temperature probes or swallow thermometer capsules), and measurement of melatonin secretion patterns has become one of the most widely used markers at the present time.

RECOMMENDED MEASURE 1: Dim Light Melatonin Onset (DLMO)[132]

Description: Melatonin is synthesized by the pineal gland and its secretion is regulated by the circadian rhythm (release occurs at night), but suppressed in the presence of light. It is the most frequently used marker of circadian rhythm since it can be measured in saliva, blood and urine.

Measurement issues: Although a frequently used measure, the lack of consistency across studies in sampling and measuring melatonin has made comparisons of results difficult. A workgroup of the Associated Professional Sleep Societies has made recommendations regarding the collection and analysis of melatonin[133]. They suggested that urine collection every 2-8 hours over 24-48 hours may be most practical in the home setting, although less precise than other approaches. Melatonin can also be measured from saliva samples collected at home or in a research or clinical setting every 30-60 min under dim lighting conditions (<30 lux), but this requires waking the subject across the night. Frequent blood sampling for melatonin through an indwelling catheter potentially allows the patient to sleep, but is the most invasive technique and must be performed in a research or clinical setting. Melatonin levels are higher in plasma than in saliva or urine, making plasma sampling the most sensitive method to estimate circadian phase, amplitude and duration of secretion. The most commonly used phase marker is dim-light melatonin onset (DMLO), usually obtained by measuring melatonin every 20-30 min for several hours prior to the normal sleep period, or dim-light melatonin offset (DLMOff), the time when melatonin levels drop at the end of the night. There is still lack of consistency as to the thresholds or methodologies used to calculate DLMO or DLMOff, however. Other issues include the effects of light exposure prior to and during sampling, and the fact that melatonin levels are low in some individuals.

RECOMMENDED MEASURE 2: Longitudinal actigraphy (acrophase, mesor, amplitude)[134,135]

Description: Actigraphy involves wearing a wristwatch-like monitor that contains a movement detector (accelerometer) and can sample and store movement data in as little as 1 second bins over weeks. Wrist actigraphy has frequently been used to estimate sleep, based on the fact that little movement occurs during sleep. It also measures activity patterns over the 24-hour day,

which are used to estimate circadian parameters such acrophase (time of peak activity), mesor (average activity), phase of the circadian rhythm and regularity of rest-activity patterns. A minimum of 1 week of data, and preferably 2 weeks, should be collected for assessment, and the monitor is typically worn on the non-dominant wrist. In clinical situations it is used to estimate sleep patterns and in the evaluation of patients with circadian rhythm disorders. It has also been used successfully to evaluate rest-activity patterns and circadian rhythms in patients with psychiatric disorders.

Measurement issues: While recommended as a measure of circadian rhythms, it is not recommended as a measure of sleep. Actigraphy cannot distinguish sleep from quiet wakefulness and, in one study, was shown to be an unreliable measure of sleep duration (based on PSG) in schizophrenia patients who tended to spend long periods lying in bed awake but not moving[136]. Methodologies for analyzing data are not standardized, and algorithms for calculation of sleep and circadian parameters vary across manufacturers. There is also variability across the instruments themselves in terms of validity and concordance with other measures of circadian rhythms.

RECOMMENDED MEASURE 3: Morningness-Eveningness Questionnaire (MEQ)[137]

Description: The Horne and Ostberg Morningness-Eveningness Questionnaire consists of 19 self-report questions that ask about bedtimes and waking times, preferred times for activities and alertness. Scores range from 16 to 86, with higher scores indicating greater morningness preference.

Measurement issues: Scoring needs to be adjusted for age, since younger people tend to express more eveningness. Scores are also not necessarily consistent across subjects from different populations or cultures. Other factors that may influence scores include work schedule, particularly shift work. Scores reflect circadian preference trait and therefore the scale cannot be used to measure change[138].

RECOMMENDED MEASURE 4: Munich Chronotype Questionnaire (MCQ)[139]

Description: The MCTQ assesses chronotype by using self-reported sleep patterns on work/school days and days off. The midpoint of the sleep period or mid-sleep on days off is used to determine the chronotype, and considered to be more of a state-like measure than the MEQ.

Measurement issues: Mid-sleep on days off is influenced by sleep debt that occurs during work/school days, particularly for late chronotypes who must awaken early for work/school. Use of an alarm clock on days off also skews the results, and cannot be used in shift workers. It is a much less widely used instrument than the MEQ[138].

Promising measures requiring further development

- Gene expression patterns[140-142]
- Pupillary light reflex[143]

Considered but not recommended

- Cortisol (too many other factors can affect data)
- Core body temperature (too invasive, melatonin fairly equivalent)

References

1. Zisner A, Beauchaine TP. Psychophysiological methods and developmental psychopathology. In: Cicchetti D, Cohen DJ, eds. *Developmental Psychopathology. Vol 2. Developmental Neuroscience*. Vol 2. Hoboken, NJ: WIley.
2. Beauchaine TP, Thayer JF. Heart rate variability as a transdiagnostic biomarker of psychopathology. *International Journal of Psychophysiology*. 2015;98(2):338-350. doi:10.1016/j.ijpsycho.2015.08.004.
3. Beauchaine TP, Gatzke-Kopp LM. Instantiating the multiple levels of analysis perspective in a program of study on externalizing behavior. *Development and Psychopathology*. 2012;24(03):1003-1018. doi:10.1017/S0954579412000508.
4. Berntson GG, Cacioppo JT, Quigley KS. Cardiac psychophysiology and autonomic space in humans: Empirical perspectives and conceptual implications. *Psychological Bulletin*. 1993;114(2):296. doi:10.1037/0033-2909.114.2.296.
5. Sherwood A. Use of impedance cardiography in cardiovascular reactivity research. In: Blascovich J, Katkin ES, eds. *Cardiovascular Reactivity to Psychological Stree and Disease*. Washington: American Psychological Association; 1993:157-199. doi:10.1037/10125-007.
6. Sherwood A, Turner JR. A conceptual and methodological overview of cardiovascular reactivity research. In: Turner JR, Sherwood A, Light KC, eds. *Individual Differences in Cardiovascular Response to Stress*. New York, NY.
7. Ritz T. Studying noninvasive indices of vagal control: The need for respiratory control and the problem of target specificity. *Biological Psychology*. 2009;80(2):158-168. doi:10.1016/j.biopsycho.2008.08.003.
8. Cipryan L, Litschmannova M. Intra-day and inter-day reliability of heart rate variability measurement. *Journal of Sports Sciences*. 2013;31(2):150-158. doi:10.1080/02640414.2012.721931.
9. Reland S, Ville NS, Wong S, Carrault G, Carré F. Reliability of heart rate variability in healthy older women at rest and during orthostatic testing. *Aging Clin Exp Res*. 2005;17(4):316-321. doi:10.1007/BF03324616.
10. Sandercock GRH, Bromley PD, Brodie DA. The reliability of short-term measurements of heart rate variability. *International Journal of Cardiology*. 2005;103(3):238-247. doi:10.1016/j.ijcard.2004.09.013.
11. Takshita Sookan AJM. Heart rate variability in physically active individuals: reliability and gender characteristics. *Cardiovascular Journal of Africa*. 2012;23(2):67. doi:10.5830/CVJA-2011-018.
12. Shader TM, Gatzke-Kopp LM, Crowell SE, et al. Quantifying respiratory sinus arrhythmia: Effects of misspecifying breathing frequencies across development. *Under Review*.
13. Fleming S, Thompson M, Stevens R, et al. Normal ranges of heart rate and respiratory rate in children from birth to 18 years of age: A systematic review of observational studies. *The Lancet*.

2011;377(9770):1011-1018. doi:10.1016/S0140-6736(10)62226-X.
14. NUNAN D, Sandercock GRH, Brodie DA. A Quantitative Systematic Review of Normal Values for Short-Term Heart Rate Variability in Healthy Adults. *Pacing and Clinical Electrophysiology*. 2010;33(11):1407-1417. doi:10.1111/j.1540-8159.2010.02841.x.
15. Beauchaine TP. Future Directions in Emotion Dysregulation and Youth Psychopathology. *Journal of Clinical Child & Adolescent Psychology*. May 2015. doi:10.1080/15374416.2015.1038827.
16. Beauchaine TP, Gatzke-Kopp L, Mead HK. Polyvagal Theory and developmental psychopathology: Emotion dysregulation and conduct problems from preschool to adolescence. *Biological Psychology*. 2007;74(2):174-184. doi:10.1016/j.biopsycho.2005.08.008.
17. Yaroslavsky I, Rottenberg J. Atypical patterns of respiratory sinus arrhythmia index an endophenotype for depression. *Development and …*. 2014.
18. Yaroslavsky I, Rottenberg J, Bylsma LM, et al. Parasympathetic nervous system activity predicts mood repair use and its effectiveness among adolescents with and without histories of major depression. *Journal of Abnormal Psychology*. 2016;125(3):323. doi:10.1037/abn0000149.
19. Neijts M, Van Lien R, Kupper N, Boomsma D, Willemsen G, Geus EJC. Heritability of cardiac vagal control in 24-h heart rate variability recordings: Influence of ceiling effects at low heart rates. *Psychophysiology*. 2014;51(10):1023-1036. doi:10.1111/psyp.12246.
20. Riese H, Muñoz LM, Hartman CA, et al. Identifying Genetic Variants for Heart Rate Variability in the Acetylcholine Pathway. *PLOS ONE*. 2014;9(11):e112476. doi:10.1371/journal.pone.0112476.
21. Su S, Lampert R, Zhao J, et al. Pleiotropy of C-Reactive Protein Gene Polymorphisms With C-Reactive Protein Levels and Heart Rate Variability in Healthy Male Twins. *The American Journal of Cardiology*. 2009;104(12):1748-1754. doi:10.1016/j.amjcard.2009.07.063.
22. Su S, Lampert R, Lee F, Bremner JD. Common genes contribute to depressive symptoms and heart rate variability: the Twins Heart Study. *… Human Genetics*. 2010.
23. Richard Jennings J, Allen B, Gianaros PJ, Thayer JF, Manuck SB. Focusing neurovisceral integration: Cognition, heart rate variability, and cerebral blood flow. *Psychophysiology*. 2015;52(2):214-224. doi:10.1111/psyp.12319.
24. Jennings JR, Sheu LK, Kuan DCH, Manuck SB, Gianaros PJ. Resting state connectivity of the medial prefrontal cortex covaries with individual differences in high-frequency heart rate variability. *Psychophysiology*. 2016;53(4):444-454. doi:10.1111/psyp.12586.
25. Thayer JF, Hansen AL, Saus-Rose E, Johnsen BH. Heart Rate Variability, Prefrontal Neural Function, and Cognitive Performance: The Neurovisceral Integration Perspective on Self-regulation, Adaptation, and Health. *ann behav med*. 2009;37(2):141-153. doi:10.1007/s12160-009-9101-z.
26. Makovac E, Meeten F, Watson DR, et al. Alterations in Amygdala-Prefrontal Functional Connectivity Account for Excessive Worry and Autonomic Dysregulation in Generalized Anxiety Disorder. *Biological Psychiatry*. October 2015. doi:10.1016/j.biopsych.2015.10.013.
27. Smith R, Allen JJB, Thayer JF, Fort C, Lane RD. Increased association over time between regional frontal lobe BOLD change magnitude and cardiac vagal control with sertraline treatment for major depression. *Psychiatry Research - Neuroimaging*. 2014;224(3):225-233. doi:10.1016/j.pscychresns.2014.08.015.
28. Beauchaine TP, Gatzke-Kopp L, Neuhaus E, Chipman J, Reid MJ, Webster-Stratton C. Sympathetic- and parasympathetic-linked cardiac function and prediction of externalizing behavior, emotion regulation, and prosocial behavior among preschoolers treated for ADHD. *Journal of Consulting and Clinical Psychology*. 2013;81(3):481. doi:10.1037/a0032302.
29. Garakani A, Martinez JM, Aaronson CJ, Voustianiouk A, Kaufmann H, Gorman JM. Effect of medication and psychotherapy on heart rate variability in panic disorder. *Depression and*

Anxiety. 2009;26(3):251-258. doi:10.1002/da.20533.
30. Posadzki P, Kuzdzal A, Lee MS, Ernst E. Yoga for Heart Rate Variability: A Systematic Review and Meta-analysis of Randomized Clinical Trials. *Applied Psychophysiology Biofeedback*. 2015;40(3):239-249. doi:10.1007/s10484-015-9291-z.
31. Reinecke A, Filippini N, Berna C, et al. Effective emotion regulation strategies improve fMRI and ECG markers of psychopathology in panic disorder: Implications for psychological treatment action. *Translational Psychiatry*. 2015;5. doi:10.1038/tp.2015.160.
32. Wang SM, Lee HK, Kweon YS, et al. Effect of emotion regulation training in patients with panic disorder: Evidenced by heart rate variability measures. *General Hospital Psychiatry*. 2016;40:68-73. doi:10.1016/j.genhosppsych.2016.01.003.
33. Clamor A, Lincoln TM, Thayer JF, Koenig J. Resting vagal activity in schizophrenia: Meta-Analysis of heart rate variability as a potential endophenotype. *British Journal of Psychiatry*. 2016;208(1):9-16. doi:10.1192/bjp.bp.114.160762.
34. Bassett D. A literature review of heart rate variability in depressive and bipolar disorders. *Australian and New Zealand Journal of Psychiatry*. 2015;50(6):511-519. doi:10.1177/0004867415622689.
35. Dawson ME, Schell AM, Filion DL. The electrodermal system. In: Cacioppo JT, Tassinary LG, Berntson GG, eds. *Handbook of Psychophysiology*. 3rd ed. New York; 2007:159-181.
36. Boucsein W. *Electrodermal Activity*. Boston, MA: Springer US; 2011. doi:10.1007/978-1-4614-1126-0_1.
37. Williams LM, Phillips ML, Brammer MJ, et al. Arousal Dissociates Amygdala and Hippocampal Fear Responses: Evidence from Simultaneous fMRI and Skin Conductance Recording. *NeuroImage*. 2001;14(5):1070-1079. doi:10.1006/nimg.2001.0904.
38. Crider A, Kremen WS, Xian H, et al. Stability, consistency, and heritability of electrodermal response lability in middle-aged male twins. *Psychophysiology*. 2004;41(4):501-509. doi:10.1111/j.1469-8986.2004.00189.x.
39. 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. doi:10.1111/psyp.12346.
40. Boucsein W, Fowles DC, Grimnes S, et al. Publication recommendations for electrodermal measurements. *Psychophysiology*. 2012;49(8):1017-1034. doi:10.1111/j.1469-8986.2012.01384.x.
41. Herpertz SC, Mueller B, Wenning B, Qunaibi M, Lichterfeld C, Herpertz-Dahlmann B. Autonomic responses in boys with externalizing disorders. *J Neural Transm*. 2003;110(10):1181-1195. doi:10.1007/s00702-003-0026-6.
42. HERPERTZ SC, VLOET T, MUELLER B, DOMES G, WILLMES K, HERPERTZ-DAHLMANN B. Similar Autonomic Responsivity in Boys With Conduct Disorder and Their Fathers. *Journal of the American Academy of Child & Adolescent Psychiatry*. 2007;46(4):535-544. doi:10.1097/chi.0b013e3180306286.
43. Isen JD, Iacono WG, Malone SM, McGue M. Examining electrodermal hyporeactivity as a marker of externalizing psychopathology: A twin study. *Psychophysiology*. 2012;49(8):1039-1048. doi:10.1111/j.1469-8986.2012.01394.x.
44. Isen JD, Iacono WG, Malone SM. Characterizing electrodermal response habituation: A latent class approach with application to psychopathology. *Psychophysiology*. 2013;50(10):954-962. doi:10.1111/psyp.12080.
45. Knott VJ, Bulmer DR. Effects of repetitive high intensity stimulation on electrodermal responsivity in male alcoholics and normal controls. *Addictive Behaviors*. 1985;10(2):181-185. doi:10.1016/0306-4603(85)90025-5.

46. Raine A, Lencz T, Bihrle S, LaCasse L, Colletti P. Reduced prefrontal gray matter volume and reduced autonomic activity in antisocial personality disorder. *Archives of General Psychiatry*. 2000;57(2):119-127. doi:10.1001/archpsyc.57.2.119.
47. Raine A. Biosocial Studies of Antisocial and Violent Behavior in Children and Adults: A Review. *J Abnorm Child Psychol*. 2002;30(4):311-326. doi:10.1023/A:1015754122318.
48. Iacono WG, Ficken JW, Beiser M. Electrodermal activation in first-episode psychotic patients and their first-degree relatives. *Psychiatry Research*. 1999;88(1):25-39. doi:10.1016/S0165-1781(99)00071-2.
49. 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. doi:10.1016/j.biopsych.2009.08.023.
50. Pole N. The Psychophysiology of Posttraumatic Stress Disorder: A Meta-Analysis. *Psychological Bulletin*. 2007;133(5):725-746. doi:10.1037/0033-2909.133.5.725.
51. Gilzenrat MS, Nieuwenhuis S, Jepma M, Cohen JD. Pupil diameter tracks changes in control state predicted by the adaptive gain theory of locus coeruleus function. *Cogn Affect Behav Neurosci*. 2010;10(2):252-269. doi:10.3758/CABN.10.2.252.
52. Koss MC. Pupillary dilation as an index of central nervous system α2-adrenoceptor activation. *Journal of Pharmacological Methods*. 1986;15(1):1-19. doi:10.1016/0160-5402(86)90002-1.
53. Hess EH, Polt JM. Pupil Size as Related to Interest Value of Visual Stimuli. *Science*. 1960;132(3423):349-350. doi:10.1126/science.132.3423.349.
54. Kahneman D, Beatty J. Pupil Diameter and Load on Memory. *Science*. 1966;154(3756):1583-1585. doi:10.1126/science.154.3756.1583.
55. Wierda SM, van Rijn H, Taatgen NA, Martens S. Pupil dilation deconvolution reveals the dynamics of attention at high temporal resolution. *PNAS*. 2012;109(22):8456-8460. doi:10.1073/pnas.1201858109.
56. Siegle GJ, Steinhauer SR, Stenger VA, Konecky R, Carter CS. Use of concurrent pupil dilation assessment to inform interpretation and analysis of fMRI data. *NeuroImage*. 2003;20(1):114-124. doi:10.1016/S1053-8119(03)00298-2.
57. Laeng B, Sirois S, Gredebäck G. Pupillometry: A Window to the Preconscious? *Perspectives on Psychological Science*. 2012;7(1):18-27. doi:10.1177/1745691611427305.
58. Farzin F, Scaggs F, Hervey C, Berry-Kravis E, Hessl D. Reliability of Eye Tracking and Pupillometry Measures in Individuals with Fragile X Syndrome. *J Autism Dev Disord*. 2011;41(11):1515-1522. doi:10.1007/s10803-011-1176-2.
59. Siegle GJ, Steinhauer SR, Friedman ES, Thompson WS, Thase ME. Remission Prognosis for Cognitive Therapy for Recurrent Depression Using the Pupil: Utility and Neural Correlates. *Biological Psychiatry*. 2011;69(8):726-733. doi:10.1016/j.biopsych.2010.12.041.
60. Anderson CJ, Colombo J, Shaddy DJ. Visual Scanning and Pupillary Responses in Young Children with Autism Spectrum Disorder. *Journal of Clinical and Experimental Neuropsychology*. February 2007. doi:10.1080/13803390500376790.
61. Martineau J, Hernandez N, Hiebel L, Roché L, Metzger A, Bonnet-Brilhault F. Can pupil size and pupil responses during visual scanning contribute to the diagnosis of autism spectrum disorder in children? *Journal of Psychiatric Research*. 2011;45(8):1077-1082. doi:10.1016/j.jpsychires.2011.01.008.
62. Price RB, Siegle GJ, Silk JS, et al. SUSTAINED NEURAL ALTERATIONS IN ANXIOUS YOUTH PERFORMING AN ATTENTIONAL BIAS TASK: A PUPILOMETRY STUDY. *Depression and Anxiety*. 2013;30(1):22-30. doi:10.1002/da.21966.
63. Silk JS Ph.D., Dahl RE M.D., Ryan ND M.D., et al. Pupillary Reactivity to Emotional Information in Child and Adolescent Depression: Links to Clinical and Ecological Measures. *American Journal of*

Psychiatry. 2007;164(12):1873-1880. doi:10.1176/appi.ajp.2007.06111816.

64. Sherwood A, Allen MT, Fahrenberg J, Kelsey RM, Lovallo WR, van Doornen LJP. Methodological Guidelines for Impedance Cardiography. *Psychophysiology*. 1990;27(1):1-23. doi:10.1111/j.1469-8986.1990.tb02171.x.
65. Sherwood A, Allen MT, Obrist PA, Langer AW. Evaluation of Beta-Adrenergic Influences on Cardiovascular and Metabolic Adjustments to Physical and Psychological Stress. *Psychophysiology*. 1986;23(1):89-104. doi:10.1111/j.1469-8986.1986.tb00602.x.
66. Mills PJ, Dimsdale JE. Sympathetic nervous system responses to psychosocial stressors. In: Turner JR, Sherwood A, Light KC, eds. *Individual Differences in Cardiovascular Response to Stress*. New York: Springer; 2013:3-32.
67. Richter M, Gendolla GHE. The heart contracts to reward: Monetary incentives and preejection period. *Psychophysiology*. 2009;46(3):451-457. doi:10.1111/j.1469-8986.2009.00795.x.
68. Brenner SL, Beauchaine TP, Sylvers PD. A comparison of psychophysiological and self-report measures of BAS and BIS activation. *Psychophysiology*. 2005;42(1):108-115. doi:10.1111/j.1469-8986.2005.00261.x.
69. Kelsey RM, Ornduff SR, Alpert BS. Reliability of cardiovascular reactivity to stress: Internal consistency. *Psychophysiology*. 2007;44(2):216-225. doi:10.1111/j.1469-8986.2007.00499.x.
70. Sherwood A, GIRDLER SS, BRAGDON EE, et al. Ten-year stability of cardiovascular responses to laboratory stressors. *Psychophysiology*. 1997;34(2):185-191. doi:10.1111/j.1469-8986.1997.tb02130.x.
71. Uchino BN, Uno D, Holt-Lunstad J, Flinders JB. Age-Related Differences in Cardiovascular Reactivity During Acute Psychological Stress in Men and Women. *J Gerontol B Psychol Sci Soc Sci*. 1999;54B(6):P339-P346. doi:10.1093/geronb/54B.6.P339.
72. Beauchaine TP, Katkin ES, Strassberg Z, Snarr J. Disinhibitory psychopathology in male adolescents: Discriminating conduct disorder from attention-deficit/hyperactivity disorder through concurrent assessment of multiple autonomic states. *Journal of Abnormal Psychology*. 2001;110(4):610. doi:10.1037/0021-843X.110.4.610.
73. Marsh P, Beauchaine TP, Williams B. Dissociation of sad facial expressions and autonomic nervous system responding in boys with disruptive behavior disorders. *Psychophysiology*. 2008;45(1):100-110. doi:10.1111/j.1469-8986.2007.00603.x.
74. Beauchaine TP, Hong J, Marsh P. Sex Differences in Autonomic Correlates of Conduct Problems and Aggression. *Journal of the American Academy of Child & Adolescent Psychiatry*. 2008;47(7):788-796. doi:10.1097/CHI.Ob013e318172ef4b.
75. Quigley KS, Stifter CA. A comparative validation of sympathetic reactivity in children and adults. *Psychophysiology*. 2006;43(4):357-365. doi:10.1111/j.1469-8986.2006.00405.x.
76. de Geus EJC, Kupper N, Boomsma DI, Snieder H. Bivariate Genetic Modeling of Cardiovascular Stress Reactivity: Does Stress Uncover Genetic Variance? *Psychosomatic Medicine*. 2007;69(4):356. doi:10.1097/PSY.0b013e318049cc2d.
77. Kupper N, Willemsen G, Boomsma DI, de Geus EJC. Heritability of Indices for Cardiac Contractility in Ambulatory Recordings. *Journal of Cardiovascular Electrophysiology*. 2006;17(8):877-883. doi:10.1111/j.1540-8167.2006.00535.x.
78. Snieder H, Harshfield GA, Barbeau P, Pollock DM, Pollock JS, Treiber FA. Dissecting the genetic architecture of the cardiovascular and renal stress response. *Biological Psychology*. 2002;61(1-2):73-95. doi:10.1016/S0301-0511(02)00053-4.
79. Buuse M. ROLE OF THE MESOLIMBIC DOPAMINE SYSTEM IN CARDIOVASCULAR HOMEOSTASIS. STIMULATION OF THE VENTRAL TEGMENTAL AREA MODULATES THE EFFECT OF VASOPRESSIN ON BLOOD PRESSURE IN CONSCIOUS RATS. *Clinical and Experimental Pharmacology and Physiology*. 1998;25(9):661-668. doi:10.1111/j.1440-1681.1998.tb02273.x.

80. Salomon K, Bylsma LM, White KE, Panaite V, Rottenberg J. Is blunted cardiovascular reactivity in depression mood-state dependent? A comparison of major depressive disorder remitted depression and healthy controls. *International Journal of Psychophysiology*. 2013;90(1):50-57. doi:10.1016/j.ijpsycho.2013.05.018.
81. Brenner SL, Beauchaine TP. Pre-ejection period reactivity and psychiatric comorbidity prospectively predict substance use initiation among middle-schoolers: A pilot study. *Psychophysiology*. 2011;48(11):1588-1596. doi:10.1111/j.1469-8986.2011.01230.x.
82. Tamaki M, Bang JW, Watanabe T, Sasaki Y. Night Watch in One Brain Hemisphere during Sleep Associated with the First-Night Effect in Humans. *Current Biology*. 2016;26(9):1190-1194. doi:10.1016/j.cub.2016.02.063.
83. COTE KA, MILNER CE, SMITH BA, et al. CNS arousal and neurobehavioral performance in a short-term sleep restriction paradigm. *Journal of Sleep Research*. 2009;18(3):291-303. doi:10.1111/j.1365-2869.2008.00733.x.
84. Basner M, Mollicone D, Dinges DF. Validity and sensitivity of a brief psychomotor vigilance test (PVT-B) to total and partial sleep deprivation. *Acta Astronautica*. 2011;69(11-12):949-959. doi:10.1016/j.actaastro.2011.07.015.
85. Christie MA, Bolortuya Y, Chen LC, McKenna JT, McCarley RW, Strecker RE. Microdialysis elevation of adenosine in the basal forebrain produces vigilance impairments in the rat psychomotor vigilance task. *Sleep*. 2008;31(10):1393-1398.
86. Drummond S, Bischoff-Grethe A. The neural basis of the psychomotor vigilance task. *SLEEP-NEW YORK …*. 2005.
87. Rétey JV, Adam M, Gottselig JM, et al. Adenosinergic mechanisms contribute to individual differences in sleep deprivation-induced changes in neurobehavioral function and brain rhythmic activity. *Journal of Neuroscience*. 2006;26(41):10472-10479. doi:10.1523/JNEUROSCI.1538-06.2006.
88. Iber C, Ancoli-Israel C, Chesson A, Quan SF. *The AASM Manual for the Scoring of Sleep and Associated Events: Rules, Terminology and Technical Specification*. 1st ed. Weschester, IL: American Academy of Sleep Medicine; 2007.
89. Krystal AD, Edinger JD. Measuring sleep quality. *Sleep Medicine*. 2008;9:S10-S17. doi:10.1016/S1389-9457(08)70011-X.
90. Silber MH, Ancoli-Israel S, Bonnet MH. The visual scoring of sleep in adults. *J Clin Sleep …*. 2007.
91. Warby SC, Wendt SL, Welinder P, et al. Sleep-spindle detection: crowdsourcing and evaluating performance of experts, non-experts and automated methods. *Nature Methods*. 2014;11(4):385-392. doi:10.1038/nmeth.2855.
92. Clawson BC, Durkin J, Aton SJ. Form and Function of Sleep Spindles across the Lifespan. *Neural Plasticity*. 2016;2016(17):1-16. doi:10.1155/2016/6936381.
93. Astori S, Wimmer RD, Prosser HM, et al. The CaV3.3 calcium channel is the major sleep spindle pacemaker in thalamus. *PNAS*. 2011;108(33):13823-13828. doi:10.1073/pnas.1105115108.
94. Manoach DS, Pan JQ, Purcell SM, Stickgold R. Reduced Sleep Spindles in Schizophrenia: A Treatable Endophenotype That Links Risk Genes to Impaired Cognition? *Biological Psychiatry*. October 2015. doi:10.1016/j.biopsych.2015.10.003.
95. Wells MF, Wimmer RD, Schmitt LI, Feng G, Halassa MM. Thalamic reticular impairment underlies attention deficit in Ptchd1Y/− mice. *Nature*. 2016;532(7597):58-63. doi:10.1038/nature17427.
96. Fogel SM, Smith CT. The function of the sleep spindle: A physiological index of intelligence and a mechanism for sleep-dependent memory consolidation. *Neuroscience & Biobehavioral Reviews*. 2011;35(5):1154-1165. doi:10.1016/j.neubiorev.2010.12.003.
97. Manoach DS, Cain MS, Vangel MG, Khurana A, Goff DC, Stickgold R. A failure of sleep-

dependent procedural learning in chronic, medicated schizophrenia. *Biological Psychiatry*. 2004;56(12):951-956. doi:10.1016/j.biopsych.2004.09.012.

98. Manoach DS, Demanuele C, Wamsley EJ, et al. Sleep spindle deficits in antipsychotic-naïve early course schizophrenia and in non-psychotic first-degree relatives. *Front Hum Neurosci*. 2014;8:344. doi:10.3389/fnhum.2014.00762.
99. Lustenberger C, Boyle MR, Alagapan S, Mellin JM, Vaughn BV, Fröhlich F. Feedback-Controlled Transcranial Alternating Current Stimulation Reveals a Functional Role of Sleep Spindles in Motor Memory Consolidation. *Current Biology*. 2016;26(16):2127-2136. doi:10.1016/j.cub.2016.06.044.
100. Mednick SC, McDevitt EA, Walsh JK, et al. The Critical Role of Sleep Spindles in Hippocampal-Dependent Memory: A Pharmacology Study. *J Neurosci*. 2013;33(10):4494-4504. doi:10.1523/JNEUROSCI.3127-12.2013.
101. Kales A, Rechtschaffen A. *A Manual of Standardized Terminology, Techniques and Scoring System for Sleep Stages of Human Subjects*.
102. Brunner DP, Dijk D-J, Tobler I, Borbély AA. Effect of partial sleep deprivation on sleep stages and EEG power spectra: evidence for non-REM and REM sleep homeostasis. *Electroencephalography and Clinical Neurophysiology*. 1990;75(6):492-499. doi:10.1016/0013-4694(90)90136-8.
103. Brunner DP, Dijk D-J, Borbély AA. Repeated partial sleep deprivation progressively changes the EEG during sleep and wakefulness. *Sleep: Journal of Sleep Research & Sleep Medicine*. February 1993.
104. Dijk D-J, Hayes B, Czeisler CA. Dynamics of electroencephalographic sleep spindles and slow wave activity in men: effect of sleep deprivation. *Brain Research*. 1993;626(1):190-199. doi:10.1016/0006-8993(93)90579-C.
105. DJ D, C C, I T, AA B. Sleep extension in humans: sleep stages, EEG power spectra and body temperature. *Sleep: Journal of Sleep Research & Sleep Medicine*. 1991;14(4):294-306.
106. Dijk D-J, Brunner DP, Beersma D, AA B. Electroencephalogram power density and slow wave sleep as a function of prior waking and circadian phase. *Sleep*. 1990.
107. Dijk DJ, Brunner DP, Borbély AA. Time course of EEG power density during long sleep in humans. *American Journal of Physiology - Regulatory, Integrative and Comparative Physiology*. 1990;258(3):R650-R661.
108. Dijk D-J, Brunner DP, Borbély AA. EEG power density during recovery sleep in the morning. *Electroencephalography and Clinical Neurophysiology*. 1991;78(3):203-214. doi:10.1016/0013-4694(91)90034-2.
109. Endo T, Schwierin B, Borbély AA, Tobler I. Selective and total sleep deprivation: effect on the sleep EEG in the rat. *Psychiatry Research*. 1997;66(2-3):97-110. doi:10.1016/S0165-1781(96)03029-6.
110. Franken P, Dijk DJ, Tobler I, Borbély AA. Sleep deprivation in rats: effects on EEG power spectra, vigilance states, and cortical temperature. *American Journal of Physiology - Regulatory, Integrative and Comparative Physiology*. 1991;261(1):R198-R208.
111. Andrew D Krystal JDE. Sleep EEG Predictors and Correlates of the Response to Cognitive Behavioral Therapy for Insomnia. *Sleep*. 2010;33(5):669.
112. Tobler I, Borbély AA. Sleep EEG in the rat as a function of prior waking. *Electroencephalography and Clinical Neurophysiology*. 1986;64(1):74-76. doi:10.1016/0013-4694(86)90044-1.
113. Werth E, Dijk DJ, Achermann P, Borbély AA. Dynamics of the sleep EEG after an early evening nap: experimental data and simulations. *American Journal of Physiology - Regulatory, Integrative and Comparative Physiology*. 1996;271(3):R501-R510.
114. Achermann P, Dijk D-J, Brunner DP, Borbély AA. A model of human sleep homeostasis based on

EEG slow-wave activity: Quantitative comparison of data and simulations. *Brain Research Bulletin*. 1993;31(1-2):97-113. doi:10.1016/0361-9230(93)90016-5.

115. Borbély AA, Baumann F, Brandeis D, Strauch I, Lehmann D. Sleep deprivation: Effect on sleep stages and EEG power density in man. *Electroencephalography and Clinical Neurophysiology*. 1981;51(5):483-493. doi:10.1016/0013-4694(81)90225-X.
116. American SOPCOT. Practice parameters for clinical use of the multiple sleep latency test and the maintenance of wakefulness test. *Sleep*. 2005.
117. D A, M B, T H, M M, R R, RB S. The clinical use of the MSLT and MWT. *Sleep: Journal of Sleep Research & Sleep Medicine*. 2005;28(1):123-144.
118. Bastien C. Validation of the Insomnia Severity Index as an outcome measure for insomnia research. *Sleep Medicine*. 2001;2(4):297-307. doi:10.1016/S1389-9457(00)00065-4.
119. Karni A, Meyer G, Rey-Hipolito C, et al. The acquisition of skilled motor performance: Fast and slow experience-driven changes in primary motor cortex. *PNAS*. 1998;95(3):861-868.
120. Walker MP, Brakefield T, Morgan A, Hobson JA, Stickgold R. Practice with Sleep Makes Perfect. *Neuron*. 2002;35(1):205-211. doi:10.1016/S0896-6273(02)00746-8.
121. Albouy G, Fogel S, Pottiez H, et al. Daytime Sleep Enhances Consolidation of the Spatial but Not Motoric Representation of Motor Sequence Memory. *PLOS ONE*. 2013;8(1):e52805. doi:10.1371/journal.pone.0052805.
122. Barakat M, Doyon J, Debas K, et al. Fast and slow spindle involvement in the consolidation of a new motor sequence. *Behavioural Brain Research*. 2011;217(1):117-121. doi:10.1016/j.bbr.2010.10.019.
123. Nishida M, Walker MP. Daytime Naps, Motor Memory Consolidation and Regionally Specific Sleep Spindles. *PLOS ONE*. 2007;2(4):e341. doi:10.1371/journal.pone.0000341.
124. Tamaki M, Huang TR, Yotsumoto Y, et al. Enhanced spontaneous oscillations in the supplementary motor area are associated with sleep-dependent offline learning of finger-tapping motor-sequence task. *Journal of Neuroscience*. 2013;33(34):13894-13902. doi:10.1523/JNEUROSCI.1198-13.2013.
125. Dresler M, Kluge M, Pawlowski M, Schüssler P, Steiger A, Genzel L. A double dissociation of memory impairments in major depression. *Journal of Psychiatric Research*. 2011;45(12):1593-1599. doi:10.1016/j.jpsychires.2011.07.015.
126. Wamsley EJ, Tucker MA, Shinn AK, et al. Reduced Sleep Spindles and Spindle Coherence in Schizophrenia: Mechanisms of Impaired Memory Consolidation? *Biological Psychiatry*. 2012;71(2):154-161. doi:10.1016/j.biopsych.2011.08.008.
127. Genzel L, Dresler M, Cornu M, et al. Medial prefrontal-hippocampal connectivity and motor memory consolidation in depression and schizophrenia. *Biological Psychiatry*. 2015;77(2):177-186. doi:10.1016/j.biopsych.2014.06.004.
128. Backhaus J, Born J, Hoeckesfeld R, Fokuhl S, Hohagen F, Junghanns K. Midlife decline in declarative memory consolidation is correlated with a decline in slow wave sleep. *Learning and Memory*. 2007;14(5):336-341. doi:10.1101/lm.470507.
129. Blatter K, Cajochen C. Circadian rhythms in cognitive performance: Methodological constraints, protocols, theoretical underpinnings. *Physiology & Behavior*. 2007;90(2-3):196-208. doi:10.1016/j.physbeh.2006.09.009.
130. Goel N, Basner M, Rao H, Dinges DF. Circadian Rhythms, Sleep Deprivation, and Human Performance. In: *Chronobiology: Biological Timing in Health and Disease*. Vol 119. Progress in Molecular Biology and Translational Science. Elsevier; 2013:155-190. doi:10.1016/B978-0-12-396971-2.00007-5.
131. Herman JH. Chronobiologic Monitoring Techniques. In: *Principles and Practice of Sleep Medicine*. Elsevier Health Sciences; 2010.

132. Burgess HJ, Wyatt JK, Park M, Fogg LF. Home circadian phase assessments with measures of compliance yield accurate dim light melatonin onsets. *Sleep*. 2015;38(6):889-897. doi:10.5665/sleep.4734.
133. Benloucif S, Burgess HJ, Klerman EB, et al. Measuring melatonin in humans. *Journal of Clinical Sleep Medicine*. 2008;4(1):66-69.
134. Briscoe S, Hardy E, Pengo MF, et al. Comparison of 7 versus 14 days wrist actigraphy monitoring in a sleep disorders clinic population. *Chronobiology International*. 2014;31(3):356-362. doi:10.3109/07420528.2013.858163.
135. Tahmasian M, Khazaie H, Golshani S, Avis KT. Clinical application of actigraphy in psychotic disorders: A systematic review. *Current Psychiatry Reports*. 2013;15(6):359. doi:10.1007/s11920-013-0359-2.
136. Manoach DS, Thakkar KN, Stroynowski E, et al. Reduced overnight consolidation of procedural learning in chronic medicated schizophrenia is related to specific sleep stages. *Journal of Psychiatric Research*. 2010;44(2):112-120. doi:10.1016/j.jpsychires.2009.06.011.
137. Horne JA, Ostberg O. A self assessment questionnaire to determine Morningness Eveningness in human circadian rhythms. *International Journal of Chronobiology*. 1976;4(2):97-110.
138. Levandovski R, Sasso E, Hidalgo MP. Chronotype: A review of the advances, limits and applicability of the main instruments used in the literature to assess human phenotype. *Trends in Psychiatry and Psychotherapy*. 2013;35(1):3-11. doi:10.1590/S2237-60892013000100002.
139. Roenneberg T, Wirz-Justice A, Merrow M. Life between clocks: Daily temporal patterns of human chronotypes. *Journal of Biological Rhythms*. 2003;18(1):80-90. doi:10.1177/0748730402239679.
140. Chen CY, Logan RW, Ma T, et al. Effects of aging on circadian patterns of gene expression in the human prefrontal cortex. *Proceedings of the National Academy of Sciences of the United States of America*. 2016;113(1):206-211. doi:10.1073/pnas.1508249112.
141. Lech K, Ackermann K, Revell VL, Lao OSCAR, Skene DJ, Kayser M. Dissecting Daily and Circadian Expression Rhythms of Clock-Controlled Genes in Human Blood. *Journal of Biological Rhythms*. 2016;31(1):68-81. doi:10.1177/0748730415611761.
142. Zhang R, Lahens NF, Ballance HI, Hughes ME, Hogenesch JB. A circadian gene expression atlas in mammals: Implications for biology and medicine. *Proceedings of the National Academy of Sciences of the United States of America*. 2014;111(45):16219-16224. doi:10.1073/pnas.1408886111.
143. Münch M, Léon L, Crippa SV, Kawasaki A. Circadian and wake-dependent effects on the pupil light reflex in response to narrow-bandwidth light pulses. *Investigative Ophthalmology and Visual Science*. 2012;53(8):4546-4555. doi:10.1167/iovs.12-9494.

APPENDIX ARS-Ia: Summary of discussion of how well each task met the proposed criteria for task evaluation.

Task: **Electrodermal conductance (a) during resting state (measure of tonic arousal); (b) in response to a task**

Construct: AROUSAL

1. How strong is the evidence that the task provides a valid measure of the RDoC construct? Very widely used test.

2. How good is the evidence about the psychometrics characteristics of the task (e.g., internal reliability, test retest, floor and ceiling effects, practice effects, availability of alternate forms, longitudinal stability)? Reasonably good for longitudinal stability and test-retest reliability for measure of tonic arousal; practice effects for repeated measurement in response to stimuli. However, insufficient data in terms of correlation with treatment response.

3. Are parameters for administering the task (e.g., number of trials, stimulus characteristics, primary dependent measure) standardized on an empirical basis? No.

4. To what extent is the task (or different versions of the task) suitable for use across lab-based studies, clinical trials (as a measure of target engagement or clinical outcome), and/or high-throughput screening settings? No standardized paradigms that have been use; has not been used in clinical outcome studies, although it has been correlated with stress response and abnormalities reported in psychiatric populations. See Peter Lang studies.

5. Is the task free from floor/ceiling effects so that it can be used across individuals with the full range of performance/impairment on the tasks? Unknown, but there are non-responders in task-response version.

6. Can the task be used (or adapted for use) with children and other special populations? Can it be used across age groups? Can it be used across different cultural settings? Yes.

7. Can the task be used as a stand-alone behavioral task? N/A

8. Is the task suitable for use in human subjects in a variety of lab environments? Is the task feasible for administration across sites? What work is needed to make the task ready for use in clinical trials?

9. Are adequate normative data available across age, gender, education, ethnicity, SES? Some normative data in healthy young adults and with respect to these other variables.

10. Is the task currently in wide usage or has its use been limited to a few research groups? Widely used, but not in relation to treatment response in psychiatric disorders.

11. Is the task sensitive to within-person change? Yes for tonic measure; repeated stimuli can possibly lead to habituation. Has been used in some biofeedback studies to show change.

12. Can the task be used with methods to interrogate brain circuitry (e.g., fMRI, EEG, etc.)? Yes

13. Can the task (or its analog) be used in animals? Is an animal version available? No.

14. Are the relationships between task performance and neural signal(s) known? Some studies correlating functional imaging with electrodermal skin response.

15. Are the relationships between task performance and clinical feature(s) known? Associated with stress/anxiety.

16. Is the task freely distributed (i.e., not copyrighted)? Yes

17. Does the task assess multiple RDoC constructs or is it specific to just one? If it assesses multiple constructs, does it allow for unambiguous conclusions about the targeted construct? Affected/assesses multiple constructs such as positive/negative valence, related to stress/anxiety.

18. If there is no existing task available for a construct, is there a task that could be modified to fit the construct? N/A

APPENDIX ARS-Ib: Summary of discussion of how well each task met the proposed criteria for task evaluation.

Task: **Psychomotor vigilance task (PVT)**

Construct: AROUSAL

1. How strong is the evidence that the task provides a valid measure of the RDoC construct? Primarily a test of alertness, which is related to arousal and affected by sleep/sleep loss, circadian rhythm, time on task.

2. How good is the evidence about the psychometrics characteristics of the task (e.g., internal reliability, test retest, floor and ceiling effects, practice effects, availability of alternate forms, longitudinal stability)? Test needs to be done multiple times during the day for each time point if not done under sleep deprivation conditions; there are ceiling effects.

3. Are parameters for administering the task (e.g., number of trials, stimulus characteristics, primary dependent measure) standardized on an empirical basis? Yes.

4. To what extent is the task (or different versions of the task) suitable for use across lab-based studies, clinical trials (as a measure of target engagement or clinical outcome), and/or high-throughput screening settings? May be used in these settings.

5. Is the task free from floor/ceiling effects so that it can be used across individuals with the full range of performance/impairment on the tasks? Ceiling effects.

6. Can the task be used (or adapted for use) with children and other special populations? Can it be used across age groups? Can it be used across different cultural settings? Yes.

7. Can the task be used as a stand-alone behavioral task?

8. Is the task suitable for use in human subjects in a variety of lab environments? Is the task feasible for administration across sites? What work is needed to make the task ready for use in clinical trials?

9. Are adequate normative data available across age, gender, education, ethnicity, SES? Yes.

10. Is the task currently in wide usage or has its use been limited to a few research groups? Moderate use.

11. Is the task sensitive to within-person change? Yes.

12. Can the task be used with methods to interrogate brain circuitry (e.g., fMRI, EEG, etc.)? Yes.

13. Can the task (or its analog) be used in animals? Is an animal version available? No.

14. Are the relationships between task performance and neural signal(s) known? No.

15. Are the relationships between task performance and clinical feature(s) known? Related to ADHD, sleep deprivation.

16. Is the task freely distributed (i.e., not copyrighted)? Yes.

17. Does the task assess multiple RDoC constructs or is it specific to just one? If it assesses multiple constructs, does it allow for unambiguous conclusions about the targeted construct? Related to attention/arousal.

APPENDIX ARS-Ic: Summary of discussion of how well each task met the proposed criteria for task evaluation.

Task: **Pupillometry in controlled context**

Construct: AROUSAL

1. How strong is the evidence that the task provides a valid measure of the RDoC construct? One of the better measures of arousal.

2. How good is the evidence about the psychometrics characteristics of the task (e.g., internal reliability, test retest, floor and ceiling effects, practice effects, availability of alternate forms, longitudinal stability)? Affected by many factors; within- and between-subject variability can be problematic. Has some ceiling and floor effects.

3. Are parameters for administering the task (e.g., number of trials, stimulus characteristics, primary dependent measure) standardized on an empirical basis? More work needed on administration and analysis standardization.

4. To what extent is the task (or different versions of the task) suitable for use across lab-based studies, clinical trials (as a measure of target engagement or clinical outcome), and/or high-throughput screening settings? Suitable for clinical trials; may be more suitable for high-throughput as technology is developed.

5. Is the task free from floor/ceiling effects so that it can be used across individuals with the full range of performance/impairment on the tasks? Reasonably good for this.

6. Can the task be used (or adapted for use) with children and other special populations? Can it be used across age groups? Can it be used across different cultural settings? Yes.

7. Can the task be used as a stand-alone behavioral task? No.

8. Is the task suitable for use in human subjects in a variety of lab environments? Is the task feasible for administration across sites? What work is needed to make the task ready for use in clinical trials? Not enough experience yet; needs more work.

9. Are adequate normative data available across age, gender, education, ethnicity, SES? No.

10. Is the task currently in wide usage or has its use been limited to a few research groups? Not widespread but use increasing.

11. Is the task sensitive to within-person change? Yes, but not extensive studies, most looking at within-subject change to sleep loss; also with the caveat of variability.

12. Can the task be used with methods to interrogate brain circuitry (e.g., fMRI, EEG, etc.)? Yes.

13. Can the task (or its analog) be used in animals? Is an animal version available? Yes; limited.

14. Are the relationships between task performance and neural signal(s) known? Yes.

15. Are the relationships between task performance and clinical feature(s) known? Yes, related to sleepiness, interest.

16. Is the task freely distributed (i.e., not copyrighted)? Yes

17. Does the task assess multiple RDoC constructs or is it specific to just one? If it assesses multiple constructs, does it allow for unambiguous conclusions about the targeted construct? Fairly tightly related to arousal, main issue is standardization.

18. If there is no existing task available for a construct, is there a task that could be modified to fit the construct?

APPENDIX ARS-Id: Summary of discussion of how well each task met the proposed criteria for task evaluation.

Task: **Heart rate variability**

Construct: AROUSAL

1. How strong is the evidence that the task provides a valid measure of the RDoC construct? Provides information on parasympathetic/sympathetic NS balance.

2. How good is the evidence about the psychometrics characteristics of the task (e.g., internal reliability, test retest, floor and ceiling effects, practice effects, availability of alternate forms, longitudinal stability)? Is affected by various factors, can be a noisy measure. In wide use, reasonable test-retest consistency.

3. Are parameters for administering the task (e.g., number of trials, stimulus characteristics, primary dependent measure) standardized on an empirical basis? Variability in administration and analysis; recommend more work to standardize.

4. To what extent is the task (or different versions of the task) suitable for use across lab-based studies, clinical trials (as a measure of target engagement or clinical outcome), and/or high-throughput screening settings? Easy to administer, more easily scalable than many other measures.

5. Is the task free from floor/ceiling effects so that it can be used across individuals with the full range of performance/impairment on the tasks?

6. Can the task be used (or adapted for use) with children and other special populations? Can it be used across age groups? Can it be used across different cultural settings?

7. Can the task be used as a stand-alone behavioral task?

8. Is the task suitable for use in human subjects in a variety of lab environments? Is the task feasible for administration across sites? What work is needed to make the task ready for use in clinical trials?

9. Are adequate normative data available across age, gender, education, ethnicity, SES? Yes.

10. Is the task currently in wide usage or has its use been limited to a few research groups? Yes.

11. Is the task sensitive to within-person change? Yes, changes in response to interventions.

12. Can the task be used with methods to interrogate brain circuitry (e.g., fMRI, EEG, etc.)? Yes.

13. Can the task (or its analog) be used in animals? Is an animal version available? Yes.

14. Are the relationships between task performance and neural signal(s) known? Some information available.

15. Are the relationships between task performance and clinical feature(s) known? Yes; associated with mood, anxiety, stress, drug effects, etc.

16. Is the task freely distributed (i.e., not copyrighted)? Yes.

17. Does the task assess multiple RDoC constructs or is it specific to just one? If it assesses multiple constructs, does it allow for unambiguous conclusions about the targeted construct? Yes, fairly specific for arousal.

18. If there is no existing task available for a construct, is there a task that could be modified to fit the construct?

APPENDIX ARS-Ie: Summary of discussion of how well each task met the proposed criteria for task evaluation.

Task: **Sleep spindles characteristics (density, amplitude, frequency, duration, topography) measured by polysomnography, particularly in stage N2 sleep.**

Construct: SLEEP-WAKEFULNESS

1. How strong is the evidence that the task provides a valid measure of the RDoC construct? Valid as a measure of thalamocortical circuitry that is involved in NREM sleep process; most information available for spindle density.

2. How good is the evidence about the psychometrics characteristics of the task (e.g., internal reliability, test retest, floor and ceiling effects, practice effects, availability of alternate forms, longitudinal stability)? Stable measure with excellent test-retest reliability.

3. Are parameters for administering the task (e.g., number of trials, stimulus characteristics, primary dependent measure) standardized on an empirical basis? EEG recording and visual spindle identification standardized, but there is some variability across labs regarding frequency definition.

4. To what extent is the task (or different versions of the task) suitable for use across lab-based studies, clinical trials (as a measure of target engagement or clinical outcome), and/or high-throughput screening settings? Depends at present on performing overnight sleep study in laboratory, so less suitable for high-throughput settings. Need to determine if nap data would be sufficient.

5. Is the task free from floor/ceiling effects so that it can be used across individuals with the full range of performance/impairment on the tasks? Yes.

6. Can the task be used (or adapted for use) with children and other special populations? Can it be used across age groups? Can it be used across different cultural settings? Yes.

7. Can the task be used as a stand-alone behavioral task? N/A.

8. Are adequate normative data available across age, gender, education, ethnicity, SES? Pending.

9. Is the task currently in wide usage or has its use been limited to a few research groups? Usage increasing.

10. Is the task sensitive to within-person change? Yes, but more trait-like.

11. Can the task be used with methods to interrogate brain circuitry (e.g., fMRI, EEG, etc.)? Yes.

12. Can the task (or its analog) be used in animals? Is an animal version available? Yes.

13. Are the relationships between task performance and neural signal(s) known? Yes.

14. Are the relationships between task performance and clinical feature(s) known? Yes.

15. Is the task freely distributed (i.e., not copyrighted)? Yes.

16. Does the task assess multiple RDoC constructs or is it specific to just one? If it assesses multiple constructs, does it allow for unambiguous conclusions about the targeted construct? Fairly specific to sleep, although associated with cognition as well as sleep.

17. If there is no existing task available for a construct, is there a task that could be modified to fit the construct? N/A

APPENDIX ARS-If: Summary of discussion of how well each task met the proposed criteria for task evaluation.

Task: **Multiple Sleep Latency Test (MSLT)**

Construct: SLEEP-WAKEFULNESS

1. How strong is the evidence that the task provides a valid measure of the RDoC construct? Strong measure of sleepiness; considered a gold standard.

2. How good is the evidence about the psychometrics characteristics of the task (e.g., internal reliability, test retest, floor and ceiling effects, practice effects, availability of alternate forms, longitudinal stability)? Limited only by "noise" related to variability.

3. Are parameters for administering the task (e.g., number of trials, stimulus characteristics, primary dependent measure) standardized on an empirical basis? Yes.

4. To what extent is the task (or different versions of the task) suitable for use across lab-based studies, clinical trials (as a measure of target engagement or clinical outcome), and/or high-throughput screening settings? Limited by need to be performed in a sleep laboratory.

5. Is the task free from floor/ceiling effects so that it can be used across individuals with the full range of performance/impairment on the tasks? Yes, ceiling effect.

6. Can the task be used (or adapted for use) with children and other special populations? Can it be used across age groups? Can it be used across different cultural settings? Yes.

7. Can the task be used as a stand-alone behavioral task? Yes.

8. Is the task suitable for use in human subjects in a variety of lab environments? Is the task feasible for administration across sites? What work is needed to make the task ready for use in clinical trials? Is used in many of these settings.

9. Are adequate normative data available across age, gender, education, ethnicity, SES? Yes.

10. Is the task currently in wide usage or has its use been limited to a few research groups? Widely used.

11. Is the task sensitive to within-person change? Yes.

12. Can the task be used with methods to interrogate brain circuitry (e.g., fMRI, EEG, etc.)? Yes.

13. Can the task (or its analog) be used in animals? Is an animal version available? No.

14. Are the relationships between task performance and neural signal(s) known? No.

15. Are the relationships between task performance and clinical feature(s) known? Yes.

16. Is the task freely distributed (i.e., not copyrighted)? Yes.

17. Does the task assess multiple RDoC constructs or is it specific to just one? If it assesses multiple constructs, does it allow for unambiguous conclusions about the targeted construct? Specific for sleepiness.

APPENDIX ARS-Ig: Summary of discussion of how well each task met the proposed criteria for task evaluation.

Task: **Polysomnographically-defined sleep architecture (sleep latency, wakefulness after sleep onset, total sleep time)**

Construct: SLEEP-WAKEFULNESS

1. How strong is the evidence that the task provides a valid measure of the RDoC construct? Gold standard.

2. How good is the evidence about the psychometrics characteristics of the task (e.g., internal reliability, test retest, floor and ceiling effects, practice effects, availability of alternate forms, longitudinal stability)? There is a floor effect on some parameters (eg, sleep latency and WASO), night-to-night variability and accommodation effects sleeping in the sleep lab.

3. Are parameters for administering the task (e.g., number of trials, stimulus characteristics, primary dependent measure) standardized on an empirical basis? Yes.

4. To what extent is the task (or different versions of the task) suitable for use across lab-based studies, clinical trials (as a measure of target engagement or clinical outcome), and/or high-throughput screening settings? Depends at present on performing overnight sleep study in laboratory, so less suitable for high-throughput settings. Need to determine if nap data would be sufficient.

5. Is the task free from floor/ceiling effects so that it can be used across individuals with the full range of performance/impairment on the tasks? Can be used across wide range of individuals but there are some floor/ceiling effects as noted above (2).

6. Can the task be used (or adapted for use) with children and other special populations? Can it be used across age groups? Can it be used across different cultural settings? Yes.

7. Can the task be used as a stand-alone behavioral task? N/A.

8. Is the task suitable for use in human subjects in a variety of lab environments? Is the task feasible for administration across sites? What work is needed to make the task ready for use in clinical trials? Suitable for use in a variety of environments and has been used extensively in multi-site clinical trials.

9. Are adequate normative data available across age, gender, education, ethnicity, SES? Yes.

10. Is the task currently in wide usage or has its use been limited to a few research groups? Widely utilized.

11. Is the task sensitive to within-person change? Yes, but limited somewhat by night-to-night variability.

12. Can the task be used with methods to interrogate brain circuitry (e.g., fMRI, EEG, etc.)? Yes.

13. Can the task (or its analog) be used in animals? Is an animal version available? Yes.

14. Are the relationships between task performance and neural signal(s) known? Sleep EEG is a neural signal.

15. Are the relationships between task performance and clinical feature(s) known? Yes.

16. Is the task freely distributed (i.e., not copyrighted)? Yes.

17. Does the task assess multiple RDoC constructs or is it specific to just one? If it assesses multiple constructs, does it allow for unambiguous conclusions about the targeted construct? Specific for this construct.

APPENDIX ARS-Ih: Summary of discussion of how well each task met the proposed criteria for task evaluation.

Task: **Slow wave measures (decline across the night as homeostatic measure, NREM average slow wave power)**

Construct: SLEEP-WAKEFULNESS

1. How strong is the evidence that the task provides a valid measure of the RDoC construct? Defining electrophysiological marker of sleep.

2. How good is the evidence about the psychometrics characteristics of the task (e.g., internal reliability, test retest, floor and ceiling effects, practice effects, availability of alternate forms, longitudinal stability)? Some night-to-night variability.

3. Are parameters for administering the task (e.g., number of trials, stimulus characteristics, primary dependent measure) standardized on an empirical basis? Some variability across labs regarding measurement.

4. To what extent is the task (or different versions of the task) suitable for use across lab-based studies, clinical trials (as a measure of target engagement or clinical outcome), and/or high-throughput screening settings? Depends at present on performing overnight sleep study in laboratory, so less suitable for high-throughput settings. Need to determine if nap data would be sufficient.

5. Is the task free from floor/ceiling effects so that it can be used across individuals with the full range of performance/impairment on the tasks? There can be floor effects in individuals without much SWA; can be confounded in elderly with generalized EEG slowing.

6. Can the task be used (or adapted for use) with children and other special populations? Can it be used across age groups? Can it be used across different cultural settings? Yes.

7. Can the task be used as a stand-alone behavioral task? No; needs to be correlated with other measures.

8. Is the task suitable for use in human subjects in a variety of lab environments? Is the task feasible for administration across sites? What work is needed to make the task ready for use in clinical trials? Easy to collect data, not all labs analyze slow waves.

9. Are adequate normative data available across age, gender, education, ethnicity, SES? Yes.

10. Is the task currently in wide usage or has its use been limited to a few research groups? Data collected widely but not analyzed in many labs.

11. Is the task sensitive to within-person change? Yes.

12. Can the task be used with methods to interrogate brain circuitry (e.g., fMRI, EEG, etc.)? Yes.

13. Can the task (or its analog) be used in animals? Is an animal version available? Yes.

14. Are the relationships between task performance and neural signal(s) known? Yes.

15. Are the relationships between task performance and clinical feature(s) known? Yes.

16. Is the task freely distributed (i.e., not copyrighted)? Yes.

17. Does the task assess multiple RDoC constructs or is it specific to just one? If it assesses multiple constructs, does it allow for unambiguous conclusions about the targeted construct? Specific to sleep.

APPENDIX ARS-Ii: Summary of discussion of how well each task met the proposed criteria for task evaluation.

Task: **Insomnia Severity Index (ISI)**

Construct: SLEEP-WAKEFULNESS

1. How strong is the evidence that the task provides a valid measure of the RDoC construct? Valid measure of insomnia and daytime consequences.

2. How good is the evidence about the psychometrics characteristics of the task (e.g., internal reliability, test retest, floor and ceiling effects, practice effects, availability of alternate forms, longitudinal stability)? Excellent.

3. Are parameters for administering the task (e.g., number of trials, stimulus characteristics, primary dependent measure) standardized on an empirical basis? Yes.

4. To what extent is the task (or different versions of the task) suitable for use across lab-based studies, clinical trials (as a measure of target engagement or clinical outcome), and/or high-throughput screening settings? Easily used; short self-administered scale.

5. Is the task free from floor/ceiling effects so that it can be used across individuals with the full range of performance/impairment on the tasks? Yes.

6. Can the task be used (or adapted for use) with children and other special populations? Can it be used across age groups? Can it be used across different cultural settings? Not validated in children, has been translated into several languages.

7. Can the task be used as a stand-alone behavioral task? Not a behavioral task.

8. Is the task suitable for use in human subjects in a variety of lab environments? Is the task feasible for administration across sites? What work is needed to make the task ready for use in clinical trials? Used widely in clinical trials.

9. Are adequate normative data available across age, gender, education, ethnicity, SES? Yes.

10. Is the task currently in wide usage or has its use been limited to a few research groups? Widely used.

11. Is the task sensitive to within-person change? Yes.

12. Can the task be used with methods to interrogate brain circuitry (e.g., fMRI, EEG, etc.)? N/A.

13. Can the task (or its analog) be used in animals? Is an animal version available? N/A.

14. Are the relationships between task performance and neural signal(s) known? Yes.

15. Are the relationships between task performance and clinical feature(s) known? Yes.

16. Is the task freely distributed (i.e., not copyrighted)? Yes.

17. Does the task assess multiple RDoC constructs or is it specific to just one? If it assesses multiple constructs, does it allow for unambiguous conclusions about the targeted construct? Measures insomnia.

APPENDIX ARS-Ij: Summary of discussion of how well each task met the proposed criteria for task evaluation.

Task: **Finger tapping Motor Sequence Task (MST)**

Construct: SLEEP-WAKEFULNESS

1. How strong is the evidence that the task provides a valid measure of the RDoC construct? Valid measure of sleep dependent memory consolidation.

2. How good is the evidence about the psychometrics characteristics of the task (e.g., internal reliability, test retest, floor and ceiling effects, practice effects, availability of alternate forms, longitudinal stability)? Good data on these parameters.

3. Are parameters for administering the task (e.g., number of trials, stimulus characteristics, primary dependent measure) standardized on an empirical basis? Yes.

4. To what extent is the task (or different versions of the task) suitable for use across lab-based studies, clinical trials (as a measure of target engagement or clinical outcome), and/or high-throughput screening settings? Needs to be done twice separated by 12 or24 hrs, which can be limiting.

5. Is the task free from floor/ceiling effects so that it can be used across individuals with the full range of performance/impairment on the tasks? Floor effect in elderly (may not show sleep-dependent learning).

6. Can the task be used (or adapted for use) with children and other special populations? Can it be used across age groups? Can it be used across different cultural settings? No norms yet in children. Age-related changes not well studied.

7. Can the task be used as a stand-alone behavioral task? Yes.

8. Is the task suitable for use in human subjects in a variety of lab environments? Is the task feasible for administration across sites? What work is needed to make the task ready for use in clinical trials? Feasible to use across sites. Used in clinical trials.

9. Are adequate normative data available across age, gender, education, ethnicity, SES? No.

10. Is the task currently in wide usage or has its use been limited to a few research groups? Most widely used probe for sleep-dependent learning, but not a lot of labs working on this.

11. Is the task sensitive to within-person change? Yes.

12. Can the task be used with methods to interrogate brain circuitry (e.g., fMRI, EEG, etc.)? Yes.

13. Can the task (or its analog) be used in animals? Is an animal version available? N/A.

14. Are the relationships between task performance and neural signal(s) known? Neuroimaging studies exist, but not many.

15. Are the relationships between task performance and clinical feature(s) known? Yes.

16. Is the task freely distributed (i.e., not copyrighted)? Yes.

17. Does the task assess multiple RDoC constructs or is it specific to just one? If it assesses multiple constructs, does it allow for unambiguous conclusions about the targeted construct? Measures sleep-dependent learning/restorative aspects of sleep fairly specifically.

APPENDIX ARS-Ik: Summary of discussion of how well each task met the proposed criteria for task evaluation.

Task: **Dim light melatonin onset (DLMO)**

Construct: CIRCADIAN RHYTHMS

1. How strong is the evidence that the task provides a valid measure of the RDoC construct? Very strong.

2. How good is the evidence about the psychometrics characteristics of the task (e.g., internal reliability, test retest, floor and ceiling effects, practice effects, availability of alternate forms, longitudinal stability)? Probably the best marker for circadian phase; caveat is that some subjects have low levels of melatonin.

3. Are parameters for administering the task (e.g., number of trials, stimulus characteristics, primary dependent measure) standardized on an empirical basis? Yes.

4. To what extent is the task (or different versions of the task) suitable for use across lab-based studies, clinical trials (as a measure of target engagement or clinical outcome), and/or high-throughput screening settings? Not ideal for high-throughput studies, otherwise useful in clinical research/trials.

5. Is the task free from floor/ceiling effects so that it can be used across individuals with the full range of performance/impairment on the tasks? Few subjects may have low levels (floor effect).

6. Can the task be used (or adapted for use) with children and other special populations? Can it be used across age groups? Can it be used across different cultural settings? Yes.

7. Can the task be used as a stand-alone behavioral task? N/A

8. Is the task suitable for use in human subjects in a variety of lab environments? Is the task feasible for administration across sites? What work is needed to make the task ready for use in clinical trials? Yes.

9. Are adequate normative data available across age, gender, education, ethnicity, SES? Some data in published studies.

10. Is the task currently in wide usage or has its use been limited to a few research groups? Used in a number of labs; samples need to be run by qualified labs.

11. Is the task sensitive to within-person change? Yes.

12. Can the task be used with methods to interrogate brain circuitry (e.g., fMRI, EEG, etc.)? N/A.

13. Can the task (or its analog) be used in animals? Is an animal version available? Yes.

14. Are the relationships between task performance and neural signal(s) known? Yes.

15. Are the relationships between task performance and clinical feature(s) known? More data needed.

16. Is the task freely distributed (i.e., not copyrighted)? Yes.

17. Does the task assess multiple RDoC constructs or is it specific to just one? If it assesses multiple constructs, does it allow for unambiguous conclusions about the targeted construct? Fairly specific for circadian phase.

18. If there is no existing task available for a construct, is there a task that could be modified to fit the construct?

APPENDIX ARS-II: Summary of discussion of how well each task met the proposed criteria for task evaluation.

Task: **Actigraphy measured longitudinally (cosinor analysis: acrophase, mesor, amplitude)**

Construct: CIRCADIAN RHYTHMS

1. How strong is the evidence that the task provides a valid measure of the RDoC construct? Fairly strong based on activity pattern.

2. How good is the evidence about the psychometrics characteristics of the task (e.g., internal reliability, test retest, floor and ceiling effects, practice effects, availability of alternate forms, longitudinal stability)? Needs optimization.

3. Are parameters for administering the task (e.g., number of trials, stimulus characteristics, primary dependent measure) standardized on an empirical basis? Yes.

4. To what extent is the task (or different versions of the task) suitable for use across lab-based studies, clinical trials (as a measure of target engagement or clinical outcome), and/or high-throughput screening settings? Easily used in clinical research, can be used for large population based studies.

5. Is the task free from floor/ceiling effects so that it can be used across individuals with the full range of performance/impairment on the tasks? Floor effect if subjects do not move around much.

6. Can the task be used (or adapted for use) with children and other special populations? Can it be used across age groups? Can it be used across different cultural settings? Yes.

7. Can the task be used as a stand-alone behavioral task? N/A

8. Is the task suitable for use in human subjects in a variety of lab environments? Is the task feasible for administration across sites? What work is needed to make the task ready for use in clinical trials? More normative data. More analytic tools to better analyze activity data (e.g., functional data analysis approach).

9. Are adequate normative data available across age, gender, education, ethnicity, SES? More needed.

10. Is the task currently in wide usage or has its use been limited to a few research groups? Fairly commonly used in circadian research.

11. Is the task sensitive to within-person change? Yes.

12. Can the task be used with methods to interrogate brain circuitry (e.g., fMRI, EEG, etc.)? N/A

13. Can the task (or its analog) be used in animals? Is an animal version available? Yes (activity monitoring commonly used).

14. Are the relationships between task performance and neural signal(s) known? Unknown.

15. Are the relationships between task performance and clinical feature(s) known? Some data in psychiatric populations.

16. Is the task freely distributed (i.e., not copyrighted)? Need to buy software.

17. Does the task assess multiple RDoC constructs or is it specific to just one? If it assesses multiple constructs, does it allow for unambiguous conclusions about the targeted construct? Data can also assess Sleep/Wakefulness and Arousal constructs.

18. If there is no existing task available for a construct, is there a task that could be modified to fit the construct?

APPENDIX ARS-Im: Summary of discussion of how well each task met the proposed criteria for task evaluation.

Task: **Munich Chronotype Questionnaire**

Construct: CIRCADIAN RHYTHMS

1. How strong is the evidence that the task provides a valid measure of the RDoC construct? Measures chronotype. In contrast to the more widely used MEQ, which is based on self-reported preferences for sleep schedule, this questionnaire determines chronotype based on reported sleep schedules.

2. How good is the evidence about the psychometrics characteristics of the task (e.g., internal reliability, test retest, floor and ceiling effects, practice effects, availability of alternate forms, longitudinal stability)? Has some limitations, such as that it is not accurate for individuals who use alarm clocks to awaken on days off from work. Not validated in shift workers.

3. Are parameters for administering the task (e.g., number of trials, stimulus characteristics, primary dependent measure) standardized on an empirical basis? Yes.

4. To what extent is the task (or different versions of the task) suitable for use across lab-based studies, clinical trials (as a measure of target engagement or clinical outcome), and/or high-throughput screening settings? May be used in all these situations.

5. Is the task free from floor/ceiling effects so that it can be used across individuals with the full range of performance/impairment on the tasks? Yes.

6. Can the task be used (or adapted for use) with children and other special populations? Can it be used across age groups? Can it be used across different cultural settings? Yes, although sociocultural factors can affect results. Pediatric version available.

7. Can the task be used as a stand-alone behavioral task? Yes.

8. Is the task suitable for use in human subjects in a variety of lab environments? Is the task feasible for administration across sites? What work is needed to make the task ready for use in clinical trials? May be used in these settings; more normative data needed.

9. Are adequate normative data available across age, gender, education, ethnicity, SES? No, more needed.

10. Is the task currently in wide usage or has its use been limited to a few research groups? Less widely used than the MEQ.

11. Is the task sensitive to within-person change? Not generally used for this purpose; chronotype a trait marker.

12. Can the task be used with methods to interrogate brain circuitry (e.g., fMRI, EEG, etc.)? N/A

13. Can the task (or its analog) be used in animals? Is an animal version available? N/A

14. Are the relationships between task performance and neural signal(s) known? Not specifically studied.

15. Are the relationships between task performance and clinical feature(s) known? Yes.

16. Is the task freely distributed (i.e., not copyrighted)? Yes, may be downloaded but website asks that permission for use be requested; no charge indicated.

17. Does the task assess multiple RDoC constructs or is it specific to just one? If it assesses multiple constructs, does it allow for unambiguous conclusions about the targeted construct? Assesses chronotype.

18. If there is no existing task available for a construct, is there a task that could be modified to fit the construct?

APPENDIX ARS-In: Summary of discussion of how well each task met the proposed criteria for task evaluation.

Task: **Horne and Ostberg Morningness-Eveningness Questionnaire (MEQ)**

Construct: CIRCADIAN RHYTHMS

1. How strong is the evidence that the task provides a valid measure of the RDoC construct? Strongly correlated with bed and waking times; strong evidence for validity.

2. How good is the evidence about the psychometrics characteristics of the task (e.g., internal reliability, test retest, floor and ceiling effects, practice effects, availability of alternate forms, longitudinal stability)? Data available confirming reliability, stability in both adult and child-adolescent versions; test-retest data more scant.

3. Are parameters for administering the task (e.g., number of trials, stimulus characteristics, primary dependent measure) standardized on an empirical basis? Yes

4. To what extent is the task (or different versions of the task) suitable for use across lab-based studies, clinical trials (as a measure of target engagement or clinical outcome), and/or high-throughput screening settings? Takes only minutes to complete, so suitable for clinical trials and high-throughput settings.

5. Is the task free from floor/ceiling effects so that it can be used across individuals with the full range of performance/impairment on the tasks? Yes.

6. Can the task be used (or adapted for use) with children and other special populations? Can it be used across age groups? Can it be used across different cultural settings? Has been translated into several languages and a pediatric version is available.

7. Can the task be used as a stand-alone behavioral task? Yes.

8. Is the task suitable for use in human subjects in a variety of lab environments? Is the task feasible for administration across sites? What work is needed to make the task ready for use in clinical trials? Yes, can be used in these settings and in clinical trials.

9. Are adequate normative data available across age, gender, education, ethnicity, SES? Some normative data (cut-off scores) available in young adults, but various factors (age, gender, socioeconomic level can affect distribution of scores).

10. Is the task currently in wide usage or has its use been limited to a few research groups? Fairly wide usage.

11. Is the task sensitive to within-person change? N/A

12. Can the task be used with methods to interrogate brain circuitry (e.g., fMRI, EEG, etc.)? N/A

13. Can the task (or its analog) be used in animals? Is an animal version available? N/A

14. Are the relationships between task performance and neural signal(s) known? N/A

15. Are the relationships between task performance and clinical feature(s) known? Correlated with biological markers of circadian phase.

16. Is the task freely distributed (i.e., not copyrighted)? Yes.

17. Does the task assess multiple RDoC constructs or is it specific to just one? If it assesses multiple constructs, does it allow for unambiguous conclusions about the targeted construct? Specific for chronotype.

APPENDIX A: RDOC MATRIX DOMAIN, CONSTRUCTS AND SUBCONSTRUCT DEFINITIONS

As defined during the initial RDoC workshops.

Arousal/Regulatory Systems: Systems responsible for generating activation of neural systems as appropriate for various contexts, and providing appropriate homeostatic regulation of such systems as energy balance and sleep.

- Arousal: Arousal is a continuum of sensitivity of the organism to stimuli, both external and internal. Arousal:
  - facilitates interaction with the environment in a context-specific manner (e.g., under conditions of threat, some stimuli must be ignored while sensitivity to and responses to others is enhanced, as exemplified in the startle reflex);
  - can be evoked by either external/environmental stimuli or internal stimuli (e.g., emotions and cognition);
  - can be modulated by the physical characteristics and motivational significance of stimuli;
  - varies along a continuum that can be quantified in any behavioral state, including wakefulness and low-arousal states including sleep, anesthesia, and coma;
  - is distinct from motivation and valence but can co-vary with intensity of motivation and valence;
  - may be associated with increased or decreased locomotor activity; and
  - can be regulated by homeostatic drives (e.g., hunger, sleep, thirst, sex).
- Circadian Rhythms: Circadian Rhythms are endogenous self-sustaining oscillations that organize the timing of biological systems to optimize physiology and behavior, and health. Circadian Rhythms:
  - are synchronized by recurring environmental cues;
  - anticipate the external environment;
  - allow effective response to challenges and opportunities in the physical and social environment;
  - modulate homeostasis within the brain and other (central/peripheral) systems, tissues and organs; and
  - are evident across levels of organization including molecules, cells, circuits, systems, organisms, and social systems.
- Sleep and wakefulness: Sleep and wakefulness are endogenous, recurring, behavioral states that reflect coordinated changes in the dynamic functional organization of the brain and that optimize physiology, behavior, and health.

Homeostatic and circadian processes regulate the propensity for wakefulness and sleep. Sleep:

- is reversible, typically characterized by postural recumbence, behavioral quiescence, and reduced responsiveness;
- has a complex architecture with predictable cycling of NREM/REM states or their developmental equivalents. NREM and REM sleep have distinct neural substrates (circuitry, transmitters, modulators) and EEG oscillatory properties
- intensity and duration is affected by homeostatic regulation;
- is affected by experiences during wakefulness;
- is evident at cellular, circuit, and system levels; and
- has restorative and transformative effects that optimize neurobehavioral functions during wakefulness.

Cognitive Systems: Systems responsible for various cognitive processes (e.g., attention, perception, memory, language, and cognitive control).

- Attention: Attention refers to a range of processes that regulate access to capacity-limited systems, such as awareness, higher perceptual processes, and motor action. The concepts of capacity limitation and competition are inherent to the concepts of selective and divided attention.
- Perception: Perception refers to the process(es) that perform computations on sensory data to construct and transform representations of the external environment, acquire information from, and make predictions about, the external world, and guide action.
- Declarative Memory: Declarative memory is the acquisition or encoding, storage and consolidation, and retrieval of representations of facts and events. Declarative memory provides the critical substrate for relational representations—i.e., for spatial, temporal, and other contextual relations among items, contributing to representations of events (episodic memory) and the integration and organization of factual knowledge (semantic memory). These representations facilitate the inferential and flexible extraction of new information from these relationships.
- Language: Language is a system of shared symbolic representations of the world, the self and abstract concepts that supports thought and communication.
- Cognitive Control: A system that modulates the operation of other cognitive and emotional systems, in the service of goal-directed behavior, when prepotent modes of responding are not adequate to meet the demands of the current context. Additionally, control processes are engaged in the case of novel contexts, where appropriate responses need to be selected from among competing alternatives.

- Working Memory: Working Memory is the active maintenance and flexible updating of goal/task relevant information (items, goals, strategies, etc.) in a form that has limited capacity and resists interference. These representations: may involve flexible binding of representations; may be characterized by the absence of external support for the internally maintained representations; and are frequently temporary, though this may be due to ongoing interference. It involves active maintenance, flexible updating, limited capacity, and interference control.

Negative Valence Systems: Systems primarily responsible for responses to aversive situations or contexts, such as:

- Responses to acute threat (Fear): Activation of the brain's defensive motivational system to promote behaviors that protect the organism from perceived danger. Normal fear involves a pattern of adaptive responses to conditioned or unconditioned threat stimuli (exteroceptive or interoceptive). Fear can involve internal representations and cognitive processing, and can be modulated by a variety of factors.
- Responses to potential harm (Anxiety): Activation of a brain system in which harm may potentially occur but is distant, ambiguous, or low/uncertain in probability, characterized by a pattern of responses such as enhanced risk assessment (vigilance). These responses to low imminence threats are qualitatively different than the high imminence threat behaviors that characterize fear.
- Responses to sustained threat: An aversive emotional state caused by prolonged (i.e., weeks to months) exposure to internal and/or external condition(s), state(s), or stimuli that are adaptive to escape or avoid. The exposure may be actual or anticipated; the changes in affect, cognition, physiology, and behavior caused by sustained threat persist in the absence of the threat, and can be differentiated from those changes evoked by acute threat.
- Frustrative non-reward: Reactions elicited in response to withdrawal/prevention of reward, i.e., by the inability to obtain positive rewards following repeated or sustained efforts.
- Loss: A state of deprivation of a motivationally significant con-specific, object, or situation. Loss may be social or non-social and may include permanent or sustained loss of shelter, behavioral control, status, loved ones, or relationships. The response to loss may be episodic (e.g., grief) or sustained.

Positive Valence Systems: Systems primarily responsible for responses to positive motivational situations or contexts, such as:

- Approach motivation: A multi-faceted construct involving mechanisms/processes that regulate the direction and maintenance of approach behavior influenced by pre-existing tendencies, learning, memory, stimulus characteristics, and deprivation states. Approach behavior can be directed toward innate or acquired cues (i.e., unconditioned vs. learned stimuli), implicit or explicit goals; it can consist of goal-directed or Pavlovian conditioned responses. Component processes include reward valuation, effort valuation/willingness to work, expectancy/reward prediction error, and action selection/decision making.
    - Reward valuation: Processes by which the probability and benefits of a prospective outcome are computed and calibrated by reference to external information, social context (e.g., group input, counterfactual comparisons), and/or prior experience. This calibration is influenced by pre-existing biases, learning, memory, stimulus characteristics, and deprivation states. Reward valuation may involve the assignment of incentive salience to stimuli.
    - Effort valuation/Willingness to work: Processes by which the cost(s) of obtaining an outcome is computed; tendency to overcome response costs to obtain a reinforcer.
    - Expectancy/Reward prediction error: A state triggered by exposure to internal or external stimuli, experiences or contexts that predict the possibility of reward. Reward expectation can alter the experience of an outcome and can influence the use of resources (e.g., cognitive resources).
    - Action selection/Preference-based decision making: Processes involving an evaluation of costs/benefits and occurring in the context of multiple potential choices being available for decision-making.
- Initial responsiveness to reward attainment: Mechanisms/processes associated with hedonic responses—as reflected in subjective experiences, behavioral responses, and/or engagement of the neural systems to a positive reinforcer—and culmination of reward seeking.
- Sustained/Longer-term responsiveness to reward attainment: Mechanisms/processes associated with the termination of reward seeking, e.g., satisfaction, satiation, regulation of consummatory behavior.
- Reward Learning: A process by which organisms acquire information about stimuli, actions, and contexts that predict positive outcomes, and by which behavior is modified when a novel reward occurs or outcomes are better than expected. Reward learning is a type of reinforcement learning, and similar processes may be involved in learning related to negative reinforcement.

- Habit: Sequential, repetitive, motor, or cognitive behaviors elicited by external or internal triggers that, once initiated, can go to completion without constant conscious oversight. Habits can be adaptive by virtue of freeing up cognitive resources. Habit formation is a frequent consequence of reward learning, but its expression can become resistant to changes in outcome value. Related behaviors could be pathological expression of a process that under normal circumstances subserves adaptive goals.

Systems for Social Processes: Systems that mediate processes to interpersonal settings of various types, including perception and interpretation of others' actions.

- Affiliation and Attachment: Affiliation is engagement in positive social interactions with other individuals. Attachment is selective affiliation as a consequence of the development of a social bond. Affiliation and Attachment are moderated by social information processing (processing of social cues) and social motivation. Affiliation is a behavioral consequence of social motivation and can manifest itself in social approach behaviors. Affiliation and Attachment require detection of and attention to social cues, as well as social learning and memory associated with the formation of relationships. Affiliation and Attachment include both the positive physiological consequences of social interactions and the behavioral and physiological consequences of disruptions to social relationships. Clinical manifestations of disruptions in Affiliation and Attachment include social withdrawal, social indifference and anhedonia, and over-attachment.
- Social Communication: A dynamic process that includes both receptive and productive aspects used for exchange of socially relevant information. Social communication is essential for the integration and maintenance of the individual in the social environment. This construct is reciprocal and interactive, and social communication abilities may appear very early in life. Social communication is distinguishable from other cognitive systems (e.g., perception, cognitive control, memory, attention) in that it particularly involves interactions with conspecifics. The underlying neural substrates of social communication evolved to support both automatic/reflexive and volitional control, including the motivation and ability to engage in social communication. Receptive aspects may be implicit or explicit; examples include affect recognition, facial recognition and characterization. Productive aspects include eye contact, expressive reciprocation, and gaze following. Although facial communication was set aside as a separate sub-construct for the purposes of identifying matrix elements, social communication typically utilizes information from several modalities, including facial, vocal, gestural, postural, and olfactory processing. Social Communication was organized into the following sub-constructs:

- Reception of Facial Communication: The capacity to perceive someone's emotional state non-verbally based on facial expressions.
- Production of Facial Communication: The capacity to convey one's emotional state non-verbally via facial expression.
- Reception of Non-Facial Communication: The capacity to perceive social and emotional information based on modalities other than facial expression, including non-verbal gestures, affective prosody, distress calling, cooing, etc.
- Production of Non-Facial Communication: The capacity to express social and emotional information based on modalities other than facial expression, including non-verbal gestures, affective prosody, distress calling, cooing, etc.

- Perception and Understanding of Self: The processes and/or representations involved in being aware of, accessing knowledge about, and/or making judgments about the self. These processes/representations can include current cognitive or emotional internal states, traits, and/or abilities, either in isolation or in relationship to others, as well as the mechanisms that support self-awareness, self-monitoring, and self-knowledge. Perception and Understanding of Self was organized into the following sub-constructs:
  - Agency: The ability to recognize one's self as the agent of one's actions and thoughts, including the recognition of one's own body/body parts.
  - Self-Knowledge: The ability to make judgments about one's current cognitive or emotional internal states, traits, and/or abilities.
- Perception and Understanding of Others: The processes and/or representations involved in being aware of, accessing knowledge about, reasoning about, and/or making judgments about other animate entities, including information about cognitive or emotional states, traits or abilities. Perception and Understanding of Others was organized into the following sub-constructs:
  - Animacy Perception: The ability to appropriately perceive that another entity is an agent (i.e., has a face, interacts contingently, and exhibits biological motion).
  - Action Perception: The ability to perceive the purpose of an action being performed by an animate entity.
  - Understanding Mental States: The ability to make judgments and/or attributions about the mental state of other animate entities that allows one to predict or interpret their behaviors. Mental state refers to intentions, beliefs, desires, and emotion

APPENDIX B: NAMHC ROSTER

National Advisory Mental Health Council

DEPARTMENT OF HEALTH AND HUMAN SERVICES
NATIONAL INSTITUTES OF HEALTH
NATIONAL INSTITUTE OF MENTAL HEALTH
NATIONAL ADVISORY MENTAL HEALTH COUNCIL
(Terms end 9/30 of designated year)

CHAIRPERSON
Bruce N. Cuthbert, Ph.D.
Acting Director
National Institute of Mental Health
Bethesda, MD

EXECUTIVE SECRETARY
Jean Noronha, Ph.D.
Director
Division of Extramural Activities
National Institute of Mental Health
Bethesda, MD

Members

Patricia A. Areán, Ph.D. (16)
Professor
Director of Targeted Treatment Development
Department of Psychiatry and Behavioral Sciences
University of Washington
Seattle, WA

Deanna M. Barch, Ph.D. (16)
Gregory B. Couch Professor of Psychiatry
Department of Psychology, Psychiatry and Radiology
Washington University
St. Louis, MO

David A. Brent, M.D. (17)
Academic Chief
Child & Adolescent Psychiatry
Endowed Chair in Suicide Studies
Professor of Psychiatry, Pediatrics and Epidemiology
Director, Services for Teens at Risk
University of Pittsburgh School of Medicine
Pittsburgh, PA

BJ Casey, Ph.D. (16)
Sackler Professor
Department of Psychiatry and Neuroscience
Sackler Institute for Developmental Psychobiology
Weill Medical College of Cornell University
New York, NY

Benjamin G. Druss, M.D., M.P.H. (18)
Rosalynn Carter Chair in Mental Health and Professor
Department of Health Policy and Management
Rollins School of Public Health
Emory University
Atlanta, GA

Hakon Heimer, M.S. (16)
Founding Editor
Schizophrenia Research Forum
Brain and Behavior Research Foundation
Providence, RI

Michael F. Hogan, Ph.D. (18)
Consultant and Advisor
Hogan Health Solutions LLC
Delmar, NY

Richard L. Huganir, Ph.D. (17)
Professor and Director
Department of Neuroscience
Investigator, Howard Hughes Medical Institute
Co-Director, Brain Science Institute
The Johns Hopkins University School of Medicine
Baltimore, MD

John H. Krystal, M.D. (19)
Robert L. McNeil, Jr. Professor of
Translational Research
Chair, Professor of Neurobiology
Chief of Psychiatry, Yale-New Haven Hospital
Department of Psychiatry
Yale University School of Medicine
New Haven, CT

Marsha M. Linehan, Ph.D. (17)
Professor and Director
Behavioral Research and Therapy Clinics
Department of Psychology
University of Washington
Seattle, WA

Maria A. Oquendo, M.D. (17)
Vice Chair for Education
Professor of Psychiatry
Department of Psychiatry
Columbia University
New York State Psychiatric Institute
New York, NY

Gene E. Robinson, Ph.D. (16)
Director, Carl R. Woese Institute for Genomic Biology
Swanlund Chair
Center for Advanced Study Professor in Entomology
And Neuroscience
University of Illinois at Urbana-Champaign
Urbana, IL

Rhonda Robinson Beale, M.D. (19)
Senior Vice President and Chief Medical Officer
Blue Cross of Idaho
Meridian, ID

Mary Jane Rotheram, Ph.D. (16)
Bat-Yaacov Professor of Child Psychiatry
And Behavioral Sciences
Director, Global Center for Children and Families
Director, Center for HIV Identification Prevention
And Treatment Services (CHIPTS)
Semel Institute and the Department of Psychiatry, University of California, Los Angeles
Los Angeles, CA

J. David Sweatt, Ph.D. (16)
Professor and Chairman
Department of Pharmacology
Vanderbilt University
Nashville, TN

Hyong Un, M.D. (17)
Head of EAP & Chief Psychiatric Officer
AETNA
Blue Bell, PA

Christopher A. Walsh, M.D. (19)
Chief, Division of Genetics and Genomics
Boston Children's Hospital
Bullard Professor of Pediatrics and Neurology
Harvard Medical School
Boston, MA

EX OFFICIO MEMBERS

Office of the Secretary, DHHS
Sylvia M. Burwell
Secretary
Department of Health and Human Services
Washington, DC

National Institutes of Health
Francis Collins, M.D., Ph.D.
Director
National Institutes of Health
Bethesda, MD

Department of Veterans Affairs
Theresa Gleason, Ph.D.
Deputy, Chief Research & Development Officer
Office of Research & Development
Department of Veterans Affairs
Washington DC

Department of Defense
John W. Davison, M.B.A., Ph.D.
Chief, Conditioned-Based Specialty Care Section
Clinical Support Division
Defense Health Agency
Department of Defense
Office of the Chief Medical Officer (OCMO)
TRICARE Management Activity, OASD (HA)
Falls Church, VA

Liaison Representative
Paolo del Vecchio, M.S.W.
Director
Center for Mental Health Services
Rockville, MD

APPENDIX C: WORKGROUP ROSTER

Workgroup Co-Chairs
Deanna M. Barch, Ph.D., Co-Chair, Washington University
Maria Oquendo, M.D., Co-Chair, Columbia University

NAMCH Members
Patricia Areán, Ph.D., University of Washington
David Brent, M.D., University of Pittsburgh School of Medicine

Workgroup Members

Arousal and Regulatory Systems
Dara Manoach, Ph.D., Chair, Harvard Medical School, Massachusetts General Hospital
Theodore Beauchaine, Ph.D., Ohio State University
Ruth Benca, M.D., Ph.D., University of Wisconsin-Madison
Andrew Krystal, M.D., Duke University

Cognitive Systems
Cameron Carter, M.D., Chair, University of California, Davis
Neal Cohen, Ph.D., University of Illinois at Urbana-Champaign
Jordan DeVylder, Ph.D., University of Maryland
Dwight Dickinson, Ph.D., J.D., National Institute of Mental Health
Damien Fair, Ph.D., PA-C, Oregon Health Sciences University
Marta Kutas, Ph.D., University of California, San Diego
Sohee Park, Ph.D., Vanderbilt University
Lucina Uddin, Ph.D., University of Miami

Negative Valence Systems
Stewart Shankman, Ph.D., Chair, University of Illinois at Chicago
Maria de las Mercedes Perez-Rodriguez, M.D., Ph.D., Mount Sinai School of Medicine
Emily Durbin, Ph.D., Michigan State University
Ian Gotlib, Ph.D., Stanford University
Sheri Johnson, Ph.D., University of California, Berkeley

Positive Valence Systems
Diego Pizzagalli, Ph.D., Chair, McLean Hospital/Harvard Medical School
Mauricio Delgado, Ph.D., Rutgers University
Paul Glimcher, Ph.D., New York University
Greg Hajcak, Ph.D., Stony Brook University
Michael Treadway, Ph.D., Emory University
Ben Yerys, Ph.D., Children's Hospital of Philadelphia

Systems for Social Processes
Kevin Pelphrey, Ph.D., Chair, George Washington University
Jed Elison, Ph.D., University of Minnesota
William P. Horan, Ph.D., University of California, Los Angeles
James Morris, Ph.D., University of Virginia
Lynn Paul, Ph.D., California Institute of Technology

NIMH RDoC Unit
Bruce Cuthbert, PhD., Acting Director of the Institute
Sarah Morris, Ph.D., Acting Director of the RDoC Unit
Dede Greenstein, Ph.D., NIMH RDoc Unit
Arina Kadam, MPH., NIMH RDoC Unit
Jenni Pacheco, Ph.D., NIMH RDoC Unit
Uma Vaidyanathan, Ph.D., NIMH RDoC Unit

APPENDIX D: WORKGROUP AGENDA

National Advisory Mental Health Council Workgroup on Tasks and Measures for RDoC

April 5, 2016

Conference Room A1/A2
Neuroscience Center
6001 Executive Boulevard
Rockville, Maryland 20852

Welcome and charge for the day

8:30 am to 8:45 am	Bruce Cuthbert, Ph.D. and Sarah Morris, Ph.D. **Location**: Conference Room A1/A2 **Key Topic**: Approval of official Workgroup Charge

Review of efforts to date

8:45 am to 9:15 am	Deanna Barch, Ph.D. and Maria Oquendo, M.D. **Location**: Conference Room A1/A2 **Key Topic:** Responses to RFI and other pre-meeting discussions

What are the relevant criteria and other issues to consider when evaluating a task for recommendation in the RDoC matrix?

9:15 am to 10:45am	Deanna Barch, Ph.D. and Maria Oquendo, M.D. **Location**: Conference Room A1/A2 **Key topic areas:** Finalize list of criteria to consider when recommending a task. Examples of

such criteria are:

- How strong is the evidence that the task provides a valid measure of the RDoC construct?
- How good is the evidence about the psychometrics characteristics of the task (e.g., internal reliability, test retest, floor and ceiling effects, practice effects)?
- Is there a version (s) of the tasks for which the parameters for administration (e.g., number of trials, stimulus characteristics, etc., primary dependent measure) have been standardized on an empirical basis?
- Can the task be used (or adapted for use) with children and other special populations? Can it be used across different cultural settings?
- To what extent is the task (or different versions of the task) suitable for use across lab-based studies, clinical trials (as a measure of target engagement or clinical outcome), and/or high-throughput screening settings?
- Does the task assess multiple RDoC constructs or is it specific to just one? If it assesses multiple constructs, does it allow for unambiguous conclusions about the targeted construct
- If there is no existing task available for a construct, is there a task that could be modified to fit the construct?

Outcome products of this meeting:

- A small set of tasks recommended for each construct of the RDoC matrix
- Answers to "criteria questions", and further information about test use, citations, and any other useful indications.
- Categorization of tasks/measures as being either:
    - Ready for "Prime Time", for at least certain age ranges or populations (with recommendations for where further validation or psychometrics are needed);
    - "Promising but in need of further development", with some recommendations of what that further development should be.
- Indication that no task currently exists and one would need to be developed

Morning break

10:45 am to 11:00 am

Parallel domain-specific subgroup meetings

11:00 am to 12:00pm

Moderators:

Dara Manoach, Ph.D. – Arousal and Regulatory Systems

7th Floor Conference #2

Cameron Carter, M.D. – Cognitive Systems
Rm 7117B
Stewart Shankman, Ph.D. – Negative Valence Systems
Conference A1
Diego Pizzagalli, Ph.D. – Positive Valence Systems
Conference A2
Kevin Pelphrey, Ph.D – Systems for Social Processes
7th Floor Conference #3

Discussion of specific tasks to be considered for recommendation for the RDoC matrix for each construct of the specified domain.

Lunch break

12:00 pm to 12:45 pm

If you have pre-ordered a lunch it will be available at noon in Conference Room A1/A2. Groups may decide to continue working over lunch, or to take a break.

Parallel domain-dpecific subgroup meetings

1:00 pm to 4:00 pm
(groups may take an afternoon break as needed)

Moderators:
Dara Manoach, Ph.D. – Arousal and Regulatory Systems
7th Floor Conference #2
Cameron Carter, M.D. – Cognitive Systems
Rm 7117B
Stewart Shankman, Ph.D. – Negative Valence Systems
Conference A1
Diego Pizzagalli, Ph.D. – Positive Valence Systems
Conference A2
Kevin Pelphrey, Ph.D – Systems for Social Processes
7th Floor Conference #3

Discussion of specific tasks that can be seen as recommendations for the RDoC matrix for each construct of the specified domain.

Challenges from the day, wrap-up and homework

4:00 pm to 5:00 pm

Deanna Barch, Ph.D. and Maria Oquendo, M.D.
Location: Conference Room A1/A2

National Advisory Mental Health Council Workgroup on Tasks and Measures for RDoC

April 6, 2016

Conference Room A1/A2
Neuroscience Center
6001 Executive Boulevard
Rockville, Maryland 20852

Welcome and charge for the day

8:30 am to 9:00 am	Deanna Barch, Ph.D. and Maria Oquendo, M.D. **Location**: Conference Room A1/A2 **Key Topic:** Discussion of challenges from yesterday, and how best to proceed.

Parallel domain-specific subgroup meetings

9:00 am to 11:30 am (groups may take a morning break as needed)	**Moderators:** Dara Manoach, Ph.D. – Arousal and Regulatory Systems 7th Floor Conference #2 Cameron Carter, M.D. – Cognitive Systems Rm 7117B Stewart Shankman, Ph.D. – Negative Valence Systems Conference A1 Diego Pizzagalli, Ph.D. – Positive Valence Systems Conference A2 Kevin Pelphrey, Ph.D – Systems for Social Processes 7th Floor Conference #3 Completion of recommendations for the domain's RDoC matrix tasks

Working lunch

11:45 am	If you have pre-ordered a lunch it will be delivered conference room A1/A2 at 11:45am.

Domain presentations

11:45 am to 2:00 pm	Deanna Barch, Ph.D. and Maria Oquendo, M.D. **Location**: Conference Room A1/A2 Each Domain group will be given 20 minutes to give a presentation about the tasks that they discussed, the recommendations that they made, and any challenges or complications that arose. Each presentation should include specific recommendations for each construct in the domain, as well as a categorization of each task mentioned as to whether it is ready for use, needs more work, or an indication that no useable task exists yet. 11:50-12:10 – Arousal and Regulatory Systems 12:10-12:30 – Cognitive Systems 12:30 – 12:50 – Negative Valence 12:50 – 1:10 – Positive Valence 1:10 – 1:30 – Social Processes 1:30-2:00 – General discussion

Outcomes from task recommendations and future directions

2:00 pm to 2:45 pm	Deanna Barch, Ph.D. and Maria Oquendo, M.D. **Location**: Conference Room A1/A2

Wrap-up

2:45 pm to 3:00 pm	Bruce Cuthbert Ph.D. and Sarah Morris, Ph.D. **Location**: Conference Room A1/A2