Realizing the Clinical Potential of Computational Psychiatry: Report from the Banbury Center Meeting, February 2019

Michael Browning[a*+], Cameron S. Carter[b], Christopher Chatham[c], Hanneke Den Ouden[d], Claire M Gillan[e], Justin T. Baker[f], Adam M. Chekroud[g], Roshan Cools[h], Peter Dayan[i], James Gold[j], Rita Z. Goldstein[k], Catherine A. Hartley[l], Adam Kepecs[m], Rebecca P. Lawson[n], Janaina Mourao-Miranda[o], Mary L. Phillips[p], Diego A. Pizzagalli[q], Albert Powers[r], David Rindskopf[s], Jonathan P Roiser[t], Katharina Schmack[u], Daniela Schiller[v], Miriam Sebold[w], Klaas Enno Stephan[x], Michael J Frank[y+], Quentin Huys[z+], Martin Paulus[zz+]

[*]Corresponding author
[+]Meeting co-organizers; alphabetical order
[a] Department of Psychiatry, University of Oxford, Oxford, UK & Oxford Health NHS Trust: michael.browning@psych.ox.ac.uk
[b] Departments of Psychiatry and Psychology, University of California at Davis
[c] Neuroscience & Rare Diseases (NRD), Roche Pharma Research & Early Development, Roche Innovation Center Basel
[d] Radboud University, Donders Institute for Brain, Cognition and Behaviour.
[e] School of Psychology, Trinity College Dublin, Dublin, Ireland
[f] Institute for Technology in Psychiatry, McLean Hospital and Department of Psychiatry, Harvard Medical School
[g] Spring Health, New York City
[h] Radboud University Medical Center, Department of Psychiatry, Nijmegen, The Netherlands & Donders Institute for Cognition, Brain and Behaviour, Centre for Cognitive Neuroimaging
[i] Max Planck Institute for Biological Cybernetics, Tübingen, Germany
[j] Maryland Psychiatric Research Center, University of Maryland School of Medicine
[k] Dept. of Psychiatry & Neuroscience, Icahn School of Medicine at Mount Sinai, New York, NY
[l] Department of Psychology, New York University
[m] Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
[n] Department of Psychology, University of Cambridge, Cambridge, UK.
[o] Centre for Medical Image Computing, Department of Computer Science and Max Planck Centre for Computational Psychiatry and Ageing Research, University College London, UK
[p] Department of Psychiatry, University of Pittsburgh, Pittsburgh, PA
[q] Department of Psychiatry, Harvard Medical School & McLean Hospital, Belmont, MA
[r] Department of Psychiatry, Yale University School of Medicine, New Haven, CT
[s] Educational Psychology, CUNY Graduate Center, New York, NY
[t] Institute of Cognitive Neuroscience, University College London, London, UK
[u] Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
[v] Dept. of Psychiatry & Neuroscience, Icahn School of Medicine at Mount Sinai, New York, NY

[w] Charité – Universitätsmedizin Berlin, Department of Psychiatry and Psychotherapy, Germany

[x] Translational Neuromodeling Unit (TNU), Institute for Biomedical Engineering, University of Zurich & ETH Zurich, Switzerland; Wellcome Centre for Human Neuroimaging, University College London. 12 Queen Square, London, WC1N, 3AR, UK; Max Planck Institute for Metabolism Research, Gleueler Strasse 50, 50931 Cologne, Germany

[y] Carney Institute for Brain Science, Dept of Cognitive, Linguistic & Psychological Sciences, Brown University, Providence, RI

[z] Division of Psychiatry and Max-Planck-UCL Centre for Computational Psychiatry and Ageing Research, University College London, UK and Translational Neuromodeling Unit (TNU), Institute for Biomedical Engineering, University of Zurich & ETH Zurich, Switzerland

[zz] Laureate Institute for Brain Research, Tulsa, OK

Computational psychiatry is an emerging field that examines phenomena in mental illness using formal techniques from computational neuroscience, mathematical psychology and machine learning (1–6). These techniques can be used in a theory-driven manner to gain insight into neural or cognitive processes and in a data-driven way to identify predictive and explanatory relationships in complex datasets. The approaches complement each other: theory-driven models can be used to infer mechanisms, and the resulting measurements can be used in data-driven approaches for prediction. Recent computational studies have successfully described and measured novel mechanisms in a range of disorders (7–11), have framed disorders in new and informative ways (12) and have identified predictors of treatment response (13,14). These methods hold the potential to improve identification of relevant clinical variables, and could be superior to classification based on traditional behavioral or neural data alone (15–18). However, these promising results have been slow to influence clinical practice or to improve patient outcomes.

In February 2019 a workshop was convened at the Banbury Centre at Cold Spring Harbor, NY. The purpose of the meeting was to identify key developments required in the practice and infrastructure of computational psychiatry research to accelerate its ability to address real world clinical problems in the near future. This report provides a summary of the conclusions of the meeting. At its core are suggestions to improve the measurement properties of computational assays through a rapid, iterative process that leverages coordinated waves of online and clinical testing, followed by deployment of the assays in innovative study designs to address clinically relevant questions. We particularly focus on theory-driven tasks but, where possible, the potential of data-driven approaches is highlighted. Finally, the report suggests that for the promise of computational psychiatry to be realized, the research environment must be developed to encourage large-scale, collaborative, interdisciplinary consortia. Given the focus of the report on questions of effective research practice in computational psychiatry, it is assumed that readers have some familiarity with the field. While we provide brief summaries and

examples of the concepts covered in the paper (Table 1, Figure 1), readers new to the field may find papers on the issue of clinical translation (3) or the general approaches used (1,2,5,15,16) useful to review beforehand.

We first summarize the need for computational assays with improved measurement properties and describe an iterative optimization and validation procedure by which such assays may be developed and deployed in clinically informative studies. We then consider broader adaptations to the research environment that may accelerate the translation of these techniques.

Computational assays for clinical applications – what is missing from current research practice?

Measurement

A key application of computational assays is the estimation of behavioral and cognitive variables that underlie clinical observations and measurements. Theory-driven approaches rely extensively on generative models i.e., formal descriptions of the underlying neural and mental processes that are believed to generate observations (see 19 for an example). Fitting generative models to observations has a number of advantages. First, it may allow identification and measurement of processes not easily captured by traditional analysis (20), second generative models may improve the validity and reliability by which a process is measured. For instance, generative models can incorporate processes that tie different features (e.g. reaction time and choice) and modalities (e.g. behavior and physiology) together in a holistic manner (21–24). They also allow artificial data to be simulated and therefore a degree of measurement optimization to occur *in silico* before the assay is deployed in practice (see Table 1 for details). However, these features do not by themselves guarantee that computational assays provide

reliable and valid measures of underlying processes. Rather, the measurement properties of an assay must be assessed and iteratively optimized (see Table 1 for a summary of computation-specific and general metrics of reliability and validity). Though there are notable exceptions (25–28) the issue of measurement in computational psychiatry has not yet attracted due attention. A principled and efficient process of assay development that optimizes measurement properties from the outset is a key outcome of the framework described below.

Deployment

Beyond questions of measurement, a second crucial factor in translating computational assays to clinical application is the deployment of the assays in studies that are able to address clinically useful questions. While cross-sectional designs can assess associations between symptoms and computational processes, they provide relatively limited information on the clinical utility of assays. Alternative study designs which test whether an assay provides predictive information useful to clinical decision-making, or the causal relationships between computationally measured processes and symptoms are likely to be especially important here. Data-driven techniques are particularly well suited to deployment in predictive studies.

Developmental pipeline for clinically useful computational assays

Here we outline a potential framework by which promising computational assays may be efficiently developed, validated and deployed to address clinically important questions (Figure 1). We note that the initial identification of candidate assays, informed by prior clinical, pre-clinical (including animal studies) and theoretical work, is crucial to this process. Rather than

describe this initial identification stage, which has been a central concern of computational psychiatry (29,30), we focus on how promising assays may most efficiently be developed.

Establishing and optimizing the measurement characteristics of novel assays

Step 1: Assay optimization

First, the measurement factors of the assay required to address a clinical question are selected and the structure of the assay altered to optimize these. Table 1 outlines important metrics. The selected factors may include both specific computational properties such as parameter identifiability (see Table 1) as well as practical features of an assay (e.g. duration to complete, complexity) and clinical validity (e.g. correlation with symptoms or treatment response). An "objective function", a mathematical formulation that combines task metrics to produce an overall measure of performance, may be constructed to reflect the specific priorities of a research project including factors to maximize (e.g. sensitivity to manipulations of key task variables, compliance) and minimize (e.g. task duration). The assay may then be optimized by iterative testing either in silico, using high-throughput online data collection (31) or in more deeply phenotyped clinical populations. Here, optimization occurs by systematically varying aspects of the assay's configuration (e.g., number of trials per condition, timing of stimulus presentation, reward incentives) in order to maximize the objective function. In some cases, this may also include hand-designed qualitative changes (e.g. to improve the task instructions used).Optimization of data-driven approaches may follow a similar trajectory with, for example, the data features being passed to a classifier that is optimized in terms of the predictive validity or the practicality of collecting the data. In effect, this step entails an expanded, recursive

piloting phase during which the measurement properties of an assay are leveraged to improve its performance.

Step 2: Latent structure validation

Although individual model parameters may underlie specific neurocognitive processes, many clinically relevant processes are likely to consist of a latent (not directly observable) structure of relations between multiple parameters (15,32). A useful step is therefore to describe this structure by collecting data from a range of assays within a single population of participants. Data-driven techniques such as clustering or theory-driven techniques such as generative modelling approaches can be used to determine the latent structure of the assays. Identified latent structures can be fed back to step 1 to inform the further development of assays, with the best performing (in terms of the metrics described in Table 1) being deployed as described below.

Deployment: Establishing the potential of assays as predictors, targets and mediators

Next, the potential clinical utility of assays is tested in proof-of-concept studies examining the predictive ability of the assay and/or the causal relationship between the process measured by the assay and clinically important outcomes such as symptoms.

Step 3a: Clinical prediction and covariation

Longitudinal observational studies may be used to assess whether an assay covaries with mental state changes or traits of interest and whether it has predictive validity, for example by predicting response to treatment (13). The ability of cohort studies to map the development of psychiatric symptoms may be enhanced by innovative study designs such as longitudinal yet

brief "natural challenge" studies (33) which make use of cohorts likely to encounter precipitative events expected to result in a change in psychiatric status (e.g., patients starting a new treatment).

Prediction will typically involve a combination of theory-driven and data-driven analysis, with data-driven analyses used to establish the most powerful predictors (13,34) and to address issues of dimensionality reduction as described for latent structure validation above.

Step 3b: Causality and treatment targets

A second route by which computational assays may impact clinical practice is if the process measured by the assay constitutes a viable treatment target. That is, treatments may be developed specifically to alter the computationally defined process. This question hinges crucially on the causal relationship between the measured process and clinically relevant outcomes such as symptoms or functioning. Causality is most efficiently addressed using experimental medicine designs which manipulate the computationally measured, process and then assess the consequences of the manipulation on intermediate or clinical outcomes (where this is not possible, quasi-experimental designs may also be useful (35)). Potential manipulations may involve pharmacological, brain stimulation, cognitive or psychotherapeutic techniques, the key issue being the ability of the intervention to engage and alter the computationally measured process.

Step 4: Clinical efficacy

Regardless of whether the goal of using a computational assay is to predict a clinical outcome or to guide the development of a novel treatment, the efficacy of computationally informed approaches must ultimately be assessed in clinical trials. Such trials may, for example, randomly assign patients to be treated according to a predictive algorithm or standard treatment, or to receive a computationally informed intervention vs. a control.

In summary, these four steps describe a general pipeline of clinical computational assay optimization designed to yield reliable and valid assays that are deployed in clinically informative study designs.

Evolution of the research environment

Computational assays can be applied to pre-existing datasets (36–38), and the sharing of datasets and analytic procedures is clearly of great importance. However, the process of computational assay development and deployment outlined above requires substantial structural resources well beyond those of individual laboratories. At the very least, this includes shared core infrastructure, particularly in the domain of information technology. It will necessitate common data structures, including meta-data relevant to measures, models and populations, and common ascertainment procedures across sites that enable individual labs to collect high-quality behavioral and clinical data and, where relevant, physiological or biological data in a universal format (39). Curation of the data will be required to ensure that it is findable, accessible, interoperable and reusable from the outset. Due to rising concerns about data security on the one hand and the need to provide scientists access to data on the other, the secure storage and aggregation of data across sites using a platform which itself may support data analysis, is likely to be essential (40).

Finally, the complexity of the human mind, the diversity of processes of clinical relevance and the range of computational theories and interventions represent a formidable intellectual challenge. It calls for a pooling of expertise and perspectives in appropriately designed multidisciplinary consortia distributed across laboratories that have a *common goal* and *share data and expertise*. Although it is beyond the scope of this paper to specify the nature and

scope of such consortia, they are likely to benefit from the inclusion of, at least, expert clinicians, experimentalists and theoreticians.

Conclusion

If computational methods are to deliver real advances for patients, we must ensure our approaches are reliable, robust, and address clinically meaningful questions. In this opinion paper we outline processes to improve the measurement properties and deployment of computational assays and highlight the importance of interdisciplinary collaboration.

Acknowledgements

Disclosures

References:

1. Huys QJM, Maia TV, Paulus MP (2016): Computational Psychiatry: From Mechanistic Insights to the Development of New Treatments. *BPS: CNNI* 1: 382–385.

2. Montague PR, Dolan RJ, Friston KJ, Dayan P (2012): Computational psychiatry. *Trends Cogn Sci* 16: 72–80.

3. Paulus MP, Huys QJM, Maia TV (2016): A Roadmap for the Development of Applied Computational Psychiatry. *BPS: CNNI* 1: 386–392.

4. Stephan KE, Binder EB, Breakspear M, Dayan P, Johnstone EC, Meyer-Lindenberg A, *et al.* (2016): Charting the landscape of priority problems in psychiatry, part 2: pathogenesis and aetiology. *Lancet Psychiatry* 3: 84–90.

5. Wang X-J, Krystal JH (2014): Computational psychiatry. *Neuron* 84: 638–654.

6. Kishida KT, King-Casas B, Montague PR (2010): Neuroeconomic approaches to mental disorders. *Neuron* 67: 543–554.

7. Browning M, Behrens TE, Jocham G, O'Reilly JX, Bishop SJ (2015): Anxious individuals have difficulty learning the causal statistics of aversive environments. *Nat Neurosci* 18: 590–596.

8. Collins AGE, Albrecht MA, Waltz JA, Gold JM, Frank MJ (2017): Interactions Among Working Memory, Reinforcement Learning, and Effort in Value-Based Choice: A New Paradigm and Selective Deficits in Schizophrenia. *Biol Psychiatry* 82: 431–439.

9. Huys QJM, Pizzagalli DA, Bogdan R, Dayan P (2013): Mapping anhedonia onto reinforcement learning: a behavioural meta-analysis. *Biol Mood Anxiety Disord* 3: 12.

10. Powers AR, Mathys C, Corlett PR (2017): Pavlovian conditioning-induced hallucinations result from overweighting of perceptual priors. *Science* 357: 596–600.

11. Lawson RP, Mathys C, Rees G (2017): Adults with autism overestimate the volatility of the sensory environment. *Nat Neurosci* 20: 1293–1299.

12. Braver TS, Barch DM, Cohen JD (1999): Cognition and control in schizophrenia: a computational model of dopamine and prefrontal function. *Biol Psychiatry* 46: 312–328.

13. Chekroud AM, Zotti RJ, Shehzad Z, Gueorguieva R, Johnson MK, Trivedi MH, *et al.* (2016): Cross-trial prediction of treatment outcome in depression: a machine learning approach. *Lancet Psychiatry* 3: 243–250.

14. Harlé KM, Stewart JL, Zhang S, Tapert SF, Yu AJ, Paulus MP (2015): Bayesian neural adjustment of inhibitory control predicts emergence of problem stimulant use. *Brain* 138: 3413–3426.

15. Wiecki TV, Poland J, Frank MJ (n.d.): Model-based cognitive neuroscience approaches to computational psychiatry clustering and classification. *Clinical Psychological Science: A Journal of the Association for Psychological Science* 3: 378–399.

16. Huys QJM, Maia TV, Frank MJ (2016): Computational psychiatry as a bridge from neuroscience to clinical applications. *Nat Neurosci* 19: 404–413.

17. Wiecki TV, Antoniades CA, Stevenson A, Kennard C, Borowsky B, Owen G, *et al.* (2016): A Computational Cognitive Biomarker for Early-Stage Huntington's Disease. *PLoS ONE* 11: e0148409.

18. Brodersen KH, Schofield TM, Leff AP, Ong CS, Lomakina EI, Buhmann JM, Stephan KE (2011): Generative embedding for model-based classification of fMRI data. *PLoS Comput Biol* 7: e1002079.

19. Behrens TEJ, Woolrich MW, Walton ME, Rushworth MFS (2007): Learning the value of information in an uncertain world. *Nat Neurosci* 10: 1214–1221.

20. Pulcu E, Browning M (2017): Affective bias as a rational response to the statistics of rewards and punishments. *Elife* 6. https://doi.org/10.7554/eLife.27879

21. Insel T, Cuthbert B, Garvey M, Heinssen R, Pine DS, Quinn K, *et al.* (2010): Research domain criteria (RDoC): toward a new classification framework for research on mental disorders. *Am J Psychiatry* 167: 748–751.

22. Hedge C, Powell G, Bompas A, Vivian-Griffiths S, Sumner P (2018): Low and variable correlation between reaction time costs and accuracy costs explained by accumulation models: Meta-analysis and simulations. *Psychol Bull* 144: 1200–1227.

23. Price RB, Brown V, Siegle GJ (2019): Computational Modeling Applied to the Dot-Probe Task Yields Improved Reliability and Mechanistic Insights. *Biol Psychiatry* 85: 606–612.

24. Kessels RPC (2019): Improving precision in neuropsychological assessment: Bridging the gap between classic paper-and-pencil tests and paradigms from cognitive neuroscience. *Clin Neuropsychol* 33: 357–368.

25. Moutoussis M, Bullmore ET, Goodyer IM, Fonagy P, Jones PB, Dolan RJ, *et al.* (2018): Change, stability, and instability in the Pavlovian guidance of behaviour from adolescence to young adulthood. *PLoS Comput Biol* 14: e1006679.

26. Shahar N, Hauser TU, Moutoussis M, Moran R, Keramati M, NSPN consortium, Dolan RJ (2019): Improving the reliability of model-based decision-making estimates in the two-stage decision task with reaction-times and drift-diffusion modeling. *PLoS Comput Biol* 15: e1006803.

27. Hedge C, Powell G, Sumner P (2018): The reliability paradox: Why robust cognitive tasks do not produce reliable individual differences. *Behav Res Methods* 50: 1166–1186.

28. Enkavi AZ, Eisenberg IW, Bissett PG, Mazza GL, MacKinnon DP, Marsch LA, Poldrack RA (2019): Large-scale analysis of test-retest reliabilities of self-regulation measures. *Proc Natl Acad Sci USA* 116: 5472–5477.

29. Maia TV, Huys QJM, Frank MJ (2017): Theory-Based Computational Psychiatry. *Biol Psychiatry* 82: 382–384.

30. Kurth-Nelson Z, O'Doherty JP, Barch DM, Denève S, Durstewitz D, Frank MJ, *et al.* (2016): Computational Approaches for Studying Mechanisms of Psychiatric Disorders. *Computational Psychiatry*. Cambridge, MA: MIT Press. Retrieved November 25, 2019, from https://oxfordindex.oup.com/view/10.7551/mitpress/9780262035422.003.0005?lang=en, //oxfordindex.oup.com:443/view/10.7551/mitpress/9780262035422.003.0005

31. Gillan CM, Daw ND (2016): Taking Psychiatry Research Online. *Neuron* 91: 19–23.

32. Poldrack RA, Yarkoni T (2016): From Brain Maps to Cognitive Ontologies: Informatics and the Search for Mental Structure. *Annu Rev Psychol* 67: 587–612.

33. Clarke P, MacLeod CM, Shirazee N (2008): Prepared for the worst: Readiness to acquire threat bias and susceptibility to elevate trait anxiety. *Emotion* 8: 47–57.

34. Calhoun VD, Lawrie SM, Mourao-Miranda J, Stephan KE (2017): Prediction of Individual Differences from Neuroimaging Data. *Neuroimage* 145: 135–136.

35. Marinescu IE, Lawlor PN, Kording KP (2018): Quasi-experimental causality in neuroscience and behavioural research. *Nat Hum Behav* 2: 891–898.

36. Etkin A, Patenaude B, Song YJC, Usherwood T, Rekshan W, Schatzberg AF, *et al.* (2015):
A cognitive-emotional biomarker for predicting remission with antidepressant
medications: a report from the iSPOT-D trial. *Neuropsychopharmacology* 40: 1332–
1342.

37. Trivedi MH, McGrath PJ, Fava M, Parsey RV, Kurian BT, Phillips ML, *et al.* (2016):
Establishing moderators and biosignatures of antidepressant response in clinical care
(EMBARC): Rationale and design. *J Psychiatr Res* 78: 11–23.

38. Auchter AM, Hernandez Mejia M, Heyser CJ, Shilling PD, Jernigan TL, Brown SA, *et al.*
(2018): A description of the ABCD organizational structure and communication
framework. *Dev Cogn Neurosci* 32: 8–15.

39. Gorgolewski KJ, Alfaro-Almagro F, Auer T, Bellec P, Capotă M, Chakravarty MM, *et al.*
(2017): BIDS apps: Improving ease of use, accessibility, and reproducibility of
neuroimaging data analysis methods. *PLoS Comput Biol* 13: e1005209.

40. Smucny J, Barch DM, Gold JM, Strauss ME, MacDonald AW, Boudewyn MA, *et al.* (2019):
Cross-diagnostic analysis of cognitive control in mental illness: Insights from the
CNTRACS consortium. *Schizophr Res*. https://doi.org/10.1016/j.schres.2019.01.018

41. Wilson RC, Collins A (2019): *Ten Simple Rules for the Computational Modeling of
Behavioral Data*. PsyArXiv. https://doi.org/10.31234/osf.io/46mbn

Table 1: Key Metrics of Reliability and Validity Relevant to Computational Measures (see also 41 for a summary)

Specific Computational Measures of Reliability	
Parameter recovery, identifiability and sensitive range	**Parameter recovery** is a process of validating parameterized generative models of behavior and/or neural data. It is performed *in silico*. A range of different parameter values are selected. These parameters are then used in the generative model to create synthetic data (at a realistic level of observation noise) which is passed back into the parameter estimation process; finally the recovered parameters are compared with the originals. The absolute difference between recovered and original parameters provides a measure of the ability of a task to estimate model parameters (if we can assume participant data can be described using a specific model), with smaller values being preferred. **Parameter identifiability** is a similar metric describing the degree to which model parameters exert distinct effects on the data and thus the degree to which differences in data can be confidently attributed to specific parameters. Parameter recovery and identifiability will generally not be constant over all parameter values (e.g. a very low "inverse temperature parameter", which will lead a model to frequently make random choices, will impair the recovery/identifiability of the *other* parameters of the model) and thus it is often useful to define the **sensitive range** of the parameters—the range of values over which parameter recovery and identifiability are achievable.
Model recovery	**Model recovery** assesses the degree to which a particular task can discriminate between different classes of generative models. This is achieved *in silico* by generating synthetic data using different models and then testing whether the process of model selection (see below) identifies the correct generative model. As for parameter recovery/identifiability, this can depend sensitively on the ranges of parameters used to generate the synthetic data.

Model selection	Where more than one model may be used to describe subject data, a process of **model selection** is used to select the "best" model. This process typically assesses the balance between the "fit" of the model (the degree to which the model can explain the data) and model complexity (i.e. its representational richness or flexibility to fit data in general). If two models explain the data similarly well, the simpler is preferred (Occam's razor). Taking into account the fit/complexity trade-off is important since models with higher complexity (e.g. with more parameters) will have higher accuracy than simpler models but may be capturing measurement-specific noise ("overfitting"). Model selection may also concern the qualitative ability of the model to recapitulate some important features of the data. While many computational studies select a single, best model for all participants and compare model parameters between participants, it is also possible to assess whether participants differ in the model which best describes their data. The finding that data from different participants are best described by different models may in itself be interesting and may be described using an hierarchical model in which a higher level selects between separate lower level models (note that, in the absence of a single model used across all participants, between subject comparison of model parameters is not straightforward).
Common Measures of Reliability	
Test-retest	The degree to which the measures of individuals within a group maintain a consistent relationship across time is assessed by **test-retest reliability**. Test-retest performance is a critical metric for tasks which are required to measure stable, trait-like, within-subject, processes, and for studies using correlational or longitudinal designs.
Split-half/interrater reliability	Other forms of reliability such as **split half reliability** or **interrater reliability** estimate measurement variability and may be useful in certain computational tasks**.**

Common Measures of Validity	
Clinical validity	Evidence for the **clinical validity** of a measure is provided by associations between it and clinically important outcomes such as symptom scores, treatment response or illness course.
Convergent/divergent validity	The degree to which a measure of a construct correlates with other measures of the same construct **(convergent validity)** and differs from measures of other constructs **(divergent validity)**. These metrics therefore provide an assessment of how certain we can be that we are measuring an underlying construct **(convergent validity)** and the degree to which our measure provides the same/different information to alternative measures **(divergent validity)**. Questions of convergent and divergent validity have largely been overlooked in computational psychiatry. As a result, it is not clear, for example, whether learning rates for positive outcomes in the plethora of available reward learning tasks measure the same thing.
Face and ecological validity	This reflects the degree to which a measurement appears to subjectively measure a process **(face validity)** and the degree to which it captures real life processes **(ecological validity)**. Computational approaches are able to decompose the components of complex processes and may therefore facilitate the development of more ecologically valid measures of complex real-life interactions.
Practical Characteristics of the Measure	
Measure duration, complexity and cost	These summarise key practical costs of a measure which are essential when considering how it may be optimised for a particular study or population.
Translational Relevance of the Measure	

Cross species translational potential of the task	Depending on the specific question being addressed, the potential for a behavioral measure to be deployed in non-human species may be relevant for measurement selection. For example, validation of the ability of a computational assay to infer physiological mechanisms may require a degree of experimental control that cannot be achieved in humans.

Figure 1: A suggested Process by which Computational Measures may be Optimized for Deployment in Clinical Studies. The stages illustrated here develop the sequential "roadmap" outlined by Paulus and colleagues (3) who framed this process using the metaphor of drug development. Here we provide more detail on how each stage may be realized (see section Establishing and Optimizing the Characteristics of Novel Assays) and, in this figure, describe examples of studies that develop a hypothetical computational task (text in italics) through stages which are similar to phase 1-3 of Paulus and colleagues' roadmap.