Cognitive Neurocomputational Task Reliability & Clinical Applications Consortium (CNTRACS)

Principal Investigators:
Deanna Barch, Washington University
Cameron Carter, University of California, Irvine
Molly Erickson, University of Chicago
James Gold, University of Maryland
Angus MacDonald, University of Minnesota

Version Date/Number: 4.1, September 23, 2026

TABLE OF CONTENTS

INTRODUCTION **3**
A1 Study Abstract 3
A2 Primary Hypotheses 3
B BACKGROUND **4**
B1 Prior Literature and Studies 4
B2 Rationale for this Study 6
C STUDY OBJECTIVES **6**
C1 Primary Aim 6
C2 Rationale for the Selection of Outcome Measures 7
D STUDY DESIGN **7**
D1 Overview or Design Summary 7
D2 Subject Selection and Withdrawal 7
2.a Inclusion Criteria *8*
2.a Exclusion Criteria *10*
2.b Ethical Considerations *10*
2.c Subject Recruitment Plans and Consent Process *10*
2.d Randomization Method and Blinding *13*
2.e Risks and Benefits *13*
E STUDY PROCEDURES **15**
E1 Screening for Eligibility 16
E2 Schedule of Measurements 16
E3 Safety and Adverse Events 20
3.a Safety and Compliance Monitoring *20*
F STATISTICAL PLAN **20**
F1 Sample Size Determination and Power 20
F2 Analysis Plan 21
F3 Statistical Methods 24
F4 Missing Outcome Data 24
G DATA HANDLING AND RECORD KEEPING **24**
G1 Confidentiality and Security 24
G2 Training 24
G3 Case Report Forms and Source Documents 24
G4 Records Retention 24
G5 Performance Monitoring 24
H STUDY ADMINISTRATION **24**
H1 Organization and Participating Centers 24
H2 Funding Source and Conflicts of Interest 25
H3 Committees 25
H4 Subject Stipends or Payments 25
H5 Study Timetable 25
I PUBLICATION PLAN **25**
J REFERENCES **25**

A. Introduction

A.1. Study Abstract

The Cognitive Neurocomputational Task Reliability and Clinical Applications in Serious Mental Illness (CNTRaCS) consortium has made significant progress in unraveling the computational and neural mechanisms of cognitive, sensory, and reward processing disruptions that contribute to functional disability across psychotic disorders. CNTRaCS has used state-of-the-art computational, psychometric, and electrophysiological measures to develop, optimize, and validate a suite of paradigms publicly available for use in a wide variety of settings, ranging from the clinic to large-scale studies. We have shown that applying computational psychiatry tools to parse cognitive mechanisms produces remarkably reliable estimates of precise parameters that allow us to dissociate and identify specific processes that are disrupted either across psychotic illnesses or specifically in schizophrenia. We will apply our previously validated approach to understanding the mechanisms underlying hallucinations and delusions in schizophrenia—positive symptoms highly associated with distress, suicidality, and violence that contribute to the economic burden associated with schizophrenia (estimated at $343.2 billion dollars in 2019). We will use a formal modeling framework to test a predictive coding model of positive symptoms that suggests that an imbalance between prior beliefs and incoming sensory evidence results in the experience of perception in the absence of sensory input and the formation of beliefs that are resistant to updating. We will tackle 7 paradigms designed to dissociate 4 constructs highly relevant to understanding the generation of hallucinations and delusions – perceptual (i.e., external sensory information), conceptual (relations, causes, events), contextual (change as a function of person or place), and social (intents and characteristics of people) beliefs. We need to optimize the psychometrics of these computational metrics to ensure that these tools are reliable, repeatable, tolerable, and not limited by ceiling and floor effects. We will utilize an open, flexible, and scalable framework to deliver these tools online, facilitating "big data" studies that examine whether the processes driving belief formation and updating are similar across the clinical to non-clinical spectrum. **Aim 1** is to test the hypothesis that model-based parameters for the measurement of belief are more precise and sensitive than standard behavioral methods in assessing deficits in psychotic disorders and have an enhanced capability to predict hallucinations and delusions in 200 individuals with psychosis (schizophrenia, schizoaffective) and 125 community controls, **Aim 2** is to measure and optimize the psychometric properties of computational parameters described in Aim 1 in a new sample of 200 individuals with psychosis and 125 healthy controls. **Aim 3** is to establish the feasibility and replicability of measurements and model-based analytic approaches outside the laboratory for assessing key components of social, perceptual, conceptual, and contextual belief formation updating in relationship to psychotic-like experiences in an online sample of 5000 community participants. **Aim 4** is to validate key model-based parameters against event-related potential (ERP) neurophysiological measures acquired using electroencephalogram (EEG).

A.2. Primary Hypotheses

Aim 1: In a sample of 200 people with psychotic symptoms (schizophrenia and schizoaffective disorder) (medicated and unmedicated) and 125 community controls, test the hypotheses 1) that model-based parameters for the measurement of belief are more precise and sensitive than standard behavioral methods in assessing deficits across psychotic disorders, and 2) determine which measures have the greatest sensitivity and specificity for the severity of hallucinations vs. delusions. We already have strong pilot data supporting the specific hypotheses to be tested in Aim 1.

Aim 2: In another sample of 200 psychotic patients and 125 healthy controls, measure and optimize the psychometric properties (test-retest reliability, internal validity, tolerability, and

absence of floor, ceiling, and practice effects) of computational parameters described in Aim 1. This is crucially important for these methods to be useful in treatment development and other research and clinical applications.

Aim 3: In a sample of 5000 community participants recruited through Prolific, establish the feasibility and replicability of measurements and model-based analytic approaches outside the laboratory for assessing key components of social, perceptual, conceptual, and contextual belief formation and updating. We will also assess their relationships to variation in psychotic-like experiences (including hallucinations, paranoia, unusual thought content, and beliefs) to measure convergent validity as well as measures of depression, anxiety, anhedonia, and substance use to understand discriminant validity.

Aim 4: Validate key model-based parameters against well-characterized neurophysiological measures acquired using EEG. We will examine previously validated Event-related potential (ERP) indices of speech perception, feedback receipt, and expectation violation as markers of these underlying constructs, which we predict will be more strongly associated with computational parameters than with standard behavioral metrics (e.g., accuracy).

B. Background

B.1. Prior Literature and Studies

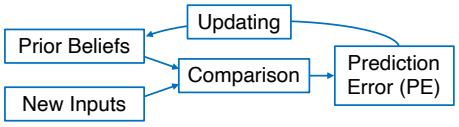
The National Institute of Mental Health (NIMH) has strongly committed to promoting computational psychiatry as a path toward a precision medicine approach to psychiatry. Much groundbreaking research has been conducted as a function of this commitment, with a major focus on cognitive and reward-related processes. The CNTRaCS consortium embraced the challenge of developing standardized versions of selected tasks that could be modeled computationally, optimizing measurement properties and providing the paradigms to the field and industry free of charge (https://sites.uci.edu/cntracs/). We have previously demonstrated that one can obtain computational parameters that are more mechanistically informative than standard behavioral metrics, yet still achieve good psychometric properties. Furthermore, these metrics relate in important ways to neural measures and thus are highly effective for advancing our understanding of computational and neurobiological mechanisms of cognitive impairment in people with serious mental illness. We now propose to apply these proven approaches to addressing another major gap in the field: namely, understanding the psychological, computational, and neural mechanisms that contribute to hallucinations and delusions.

Hallucinations and delusions are central components of psychotic disorders. They are major drivers of distress and impairment and are strongly associated with suicidality[1] and violent behavior[2] (both perpetration and victimization) across diagnostic boundaries. They can impede people's ability to form and maintain strong interpersonal relationships and to achieve their occupational goals, thereby often contributing to lifelong disability. Until recently, however, there has been little rigorous, quantitative work on the specific psychological, computational, or neural mechanisms that contribute to the formation and maintenance of hallucinations and delusions. This is partly because they are difficult to model in animal paradigms and challenging to precisely quantify in humans. However, advances in the development of neurocomputational models and novel paradigms by members of the CNTRaCS consortium have led to significant progress in identifying the specific systems that underlie hallucinations and delusions, setting the stage for treatment-development research. However, treatment research that targets these systems requires precise, well-validated, and psychometrically sound measures of these systems. The goal of the proposed project is to refine existing experimental paradigms so that they are suitable for large-scale treatment studies of hallucinations and delusions, to determine which paradigms

and computational metrics are most informative in regard to testing theoretical models of hallucinations and delusions, and to demonstrate that the resulting computational metrics have the necessary psychometric properties to serve as outcome measures in these studies. Addressing the lack of information about hallucinations and delusions is critical given the major impact on quality of life for those suffering from psychosis, and given that the economic burden of schizophrenia in 2019 was estimated to be $343.2 billion dollars[3].

We will test several different experimental paradigms, all of which are informed by a generalization of the widely used **predictive coding model** of perception and belief formation (Fig. 1). In this model, an individual has a set of prior beliefs about the state of the world (including themselves and other people), and these are compared with new inputs from the environment. If the inputs are consistent with the prior beliefs, then the beliefs remain unchanged. If the inputs do not match the prior beliefs, then a prediction error is generated, which leads to an updating of the person's beliefs about the world. The beliefs and inputs can be perceptual (e.g., the color of a leaf) or they can be conceptual (e.g., the likelihood that taking a medication will lead to symptom relief). Many important beliefs are social in nature (e.g., beliefs about whether people are trustworthy), whereas others are not (e.g., beliefs about the weather). These different beliefs vary in their sensitivity to new evidence; sometimes incorporating and other times ignoring new evidence, depending on the belief content (perceptual, conceptual, social), and the size and reliability of the prediction error that the evidence generates[4].

Figure 1. Basic predictive coding model



Versions of the predictive coding model have become an increasingly important framework for conceptualizing hallucinations and delusions,[5-12] and are also amenable to study in animal model[13]. For example, an overly strong prior belief can lead to a hallucination (perceiving that a voice is present in the absence of an actual voice) or to the maintenance of a delusion (e.g., continuing to believe that a listening device has been implanted in your head despite overwhelming evidence to the contrary). Similarly, a strong belief that the world is unstable could cause someone to disregard inputs as irrelevant for predicting the future (e.g., ignoring evidence that a medication has been effective in the past). Several experimental paradigms and computational modeling frameworks have been developed in the context of the predictive coding model—by members of the CNTRaCS team and others—that have shown promise for quantifying the specific systems and processes that underlie hallucinations and delusions[4, 12-22]. Although these paradigms and modeling approaches share a common high-level framework, they vary in terms of whether they mainly explain hallucinations or delusions, whether they focus on perceptual or conceptual information, whether they focus on the stability or instability of the environment, and whether they focus on social or nonsocial information. It is not clear whether these differences allow them to isolate distinctly different sources of dysfunction or whether they are all tapping into the same underlying pathology. Further, although we have evidence that many of these tasks yield computational parameters associated with hallucinations and/or delusions, it is not clear which one(s) provides the clearest clinical signal. Moreover, there has been no systematic effort to assess predictive coding across both hallucinations and delusions, both perceptual and conceptual tasks, and both social and nonsocial situations, so the field lacks the data needed to even begin building a single integrated model of the role of predictive coding deficits in psychosis. The goal of the proposed research is to collect the needed data from a broad set of patients and controls and to determine the tasks best suited for use in future treatment-oriented research in terms of both psychometric properties (e.g., test-retest reliability) and clinical signal (e.g., ability to predict scores on measures of hallucinations and delusions). A key element of our approach is using computational parameters that isolate components of task performance and comparing them with traditional noncomputational performance metrics. Further, collecting

all these data simultaneously will allow us to incorporate hierarchical modeling jointly across tasks and symptoms, enabling us to estimate their covariance and understand how computational parameters in one domain relate to those in another. This will provide the data needed to begin building an integrated model.

We have selected tasks with prior evidence of a clinical signal (e.g., correlations with symptom scores), including paradigms that vary in terms of perceptual vs. conceptual information content, social vs. nonsocial judgments, and whether they emphasize the (in)stability of the environment. To provide timely results, we have selected paradigms that already demonstrate a clinical signal (e.g., correlations with symptom scores) rather than inventing new tasks, but which have never been used together in an integrated way to generate a more complete model of the relationships between specific aspects of predictive coding deficits and specific symptom domains. In Table 1, we outline the specific behavioral metrics for each task, the computational metrics to which these would be compared, and the specific *a priori* theoretically selected ERP measures we will use to examine whether they are more closely linked to biological signals than are the standard behavioral metrics, and to help arbitrate between competing explanations. By collecting data across paradigms in a common set of patients and controls, as well as a very large online sample of community participants, we can assess whether these task dimensions are important for characterizing unique aspects of psychotic symptoms and determine which paradigms provide the strongest and most reliable clinically relevant signals for future treatment studies.

B.2. Rationale for this Study

We will use these tasks to address the following specific questions. First, do hallucinations emerge from overly strong top-down priors[14], or are they driven by prediction errors with noisy or unreliable precision[23]? Second, delusions often include beliefs about other agents. Thus, are delusions more specifically related to social beliefs, or are they equally related to other belief domains? Third, are disrupted perceptual beliefs more strongly associated with hallucinations, whereas disrupted conceptual beliefs are more strongly associated with delusions? Fourth, does instability in the world elicit beliefs about other agents that can become delusional? Fifth, can relatively simple tasks reveal predictive coding abnormalities just as well as more complex tasks? If so, simpler tasks will be easier to implement in large-scale studies, clinical settings, and translated to animal models. In Section 3.1, we describe each of the tasks and preliminary data, followed by our proposed approach for using computational modeling of task performance and EEG-based biomarkers to address these key open questions. We confront a central tension in this literature: delusions and hallucinations may be driven by different mechanisms[24]. They are correlated clinically, but not perfectly[25], and have different trajectories over time[26] and responses to treatment[27]. Examining tasks that engage different types of beliefs (perceptual, conceptual, social), with different contents, and computational models of prediction error, prior-weighting, and updating, will delineate the distinctions between delusions and hallucinations. Doing so is crucial to designing more effective treatments for these debilitating symptoms.

C. Study Objectives

C.1. Primary Aims

Aim 1: In a sample of 200 people with psychotic symptoms (schizophrenia and schizoaffective disorder) (medicated and unmedicated) and 125 community controls, test the hypotheses 1) that model-based parameters for the measurement of belief are more precise and sensitive than standard behavioral methods in assessing deficits across psychotic disorders, and 2) determine which measures have the greatest sensitivity and specificity for the severity of hallucinations vs. delusions. We already have strong pilot data supporting the specific hypotheses to be tested in Aim 1.

Aim 2: In another sample of 200 psychotic patients and 125 healthy controls, measure and optimize the psychometric properties (test-retest reliability, internal validity, tolerability, and absence of floor, ceiling, and practice effects) of computational parameters described in Aim 1. This is crucially important for these methods to be useful in treatment development and other research and clinical applications.
Aim 3: In a sample of 5000 community participants recruited through Prolific, establish the feasibility and replicability of measurements and model-based analytic approaches outside the laboratory for assessing key components of social, perceptual, conceptual, and contextual belief formation and updating. We will also assess their relationships to variation in psychotic-like experiences (including hallucinations, paranoia, unusual thought content, and beliefs) to measure convergent validity as well as measures of depression, anxiety, anhedonia, and substance use to understand discriminant validity.
Aim 4: Validate key model-based parameters against well-characterized neurophysiological measures acquired using EEG. We will examine previously validated ERP indices of speech perception, feedback receipt, and expectation violation as markers of these underlying constructs, which we predict will be more strongly associated with computational parameters than with standard behavioral metrics (e.g., accuracy).

C.2. Rationale for the Selection of Outcome Measures

The outcome measures were selected based on the theoretical and empirical justification provided in section B1.

D. Study Design

D.1. Overview or Design Summary

In **Study 1** we will test the ability of *each* of the seven paradigms to differentiate among participants with schizophrenia (SZ)/schizoaffective disorder (SZA) and matched healthy controls (HC) and to predict individual differences in clinical measures of hallucinations and delusions. We will also directly *compare* the tasks in terms of predictive ability and psychometric properties and determine whether the tasks share variance or explain unique variance. This will include determining whether hallucinations and delusions are best explained by separate underlying processes or by a single process. We will also test whether the model-based parameters from these studies are more precise and sensitive than standard behavioral methods for assessing deficits across psychotic disorders and for predicting hallucinations and delusions (Aim 1). We will also validate key model-based parameters against well-characterized neurophysiological measures using ERP (Aim 4). The data from Study 1 will also be used to optimize (e.g., length, floor, and ceiling effects) tasks for Study 2. **Study 2** will recruit a new sample of SZ/SZA and HC and use optimized tasks to replicate the behavioral findings from Study 1 and to assess additional psychometric properties (test-retest reliability, practice effects; Aim 2). For both studies, we will obtain data on: A) clinical symptoms, with detailed assessment of hallucinations and delusions; B) community functioning; and C) functional capacity, allowing us to examine clinical validity. **Study 3** will recruit 5,000 online community participants using Prolific. This will allow us to establish the feasibility of using these tasks on a large scale outside the laboratory, permitting assessment of generalization and the impact of measurement invariance across demographic variables. See Timeline documents for specifics of timeline for each study.

D.2. Subject Selection and Withdrawal

Study 1 Pilot Phase: Participants will be up to 2500 pilot individuals, ages 18-65, recruited either on line (through Mturk or Prolific) or tested in person at each site. The in-person participants will be medicated and unmedicated Schizophrenia/Schizoaffective disorder (SZ/SZA), and community controls. The online participants will be age 18-65 with no other inclusion or exclusion criteria.

Study 1 Behavioral Pilot: Participants will be up to 2500 pilot individuals, ages 18-65, recruited and tested in person at each site. Participants will be medicated and unmedicated Schizophrenia/Schizoaffective disorder (SZ/SZA) and community controls.

Study 1: Participants will be medicated and unmedicated Schizophrenia/Schizoaffective disorder (SZ/SZA) as well as healthy community controls (HC). Across the five sites, we will recruit and complete testing with 200 SZ and 125 HC, with an additional 50 SZ/SZA and 50 HC recruited to cover dropouts or other reasons for non-participation.

Study 2: Participants will be medicated Schizophrenia/Schizoaffective disorder (SZ/SZA) as well as healthy community controls (HC). Across the five sites, we will recruit and complete testing with 200 SZ and 125 HC, with an additional 50 SZ and 50 CON recruited to cover dropouts or other reasons for non-participation.

Study 3: Participants will be 5,000 individuals recruited online through MTurk, Prolific, or a similar online platform, matched to the 2016 US census on sex, age, race, ethnicity, education, employment status, and household income. To ensure that we obtain 5000 with complete testing, we will recruit an additional 1000 individuals.

D.2.a. Inclusion Criteria

Study 1 Pilot Phase:
1) Online testing: age 18 to 65;

2) On Site testing: General Inclusion: 1) age 18 to 65; 2) no substance disorder in past three months (confirmed with saliva drug screens, breathalyzers and saliva swab); and 3) no clinically significant head injury or neurological disease. We will also administer the Wechsler Test of Adult Reading (WTAR) as a measure of premorbid IQ [28]. As in our prior studies, we will exclude participants with a score lower than 16, as this is equivalent to an estimated premorbid IQ of less than 70. This will help to ensure that participants are capable of understanding the basic task instructions. We will ensure that the groups are similar in terms of key demographics that could influence task performance (parental socio-economic scale (SES), age, gender).

SZ/SZA Specific: 1) no medication changes in prior month or anticipated in the next month (or not taking medications, see below), and 2) outpatient or partial hospital. Schizophrenia/Schizoaffective Disorder (SZ/SZA): Diagnostic and Statistical Manual of Mental Disorders, 5th Edition (DSM-5) diagnosis of schizophrenia or schizoaffective.

Unmedicated: Additional inclusion/exclusion criteria for the unmedicated individuals are: 1) not in acute exacerbation that would require immediate hospitalization or treatment; 3) self-reported medication non-compliance for at least the four previous weeks or never having started medications. Unmedicated means antipsychotic-free for schizophrenia and schizoaffective.

Study 1 Behavioral Pilot: Participants will be up to 2500 pilot individuals, ages 18-65, recruited and tested in person at each site. Participants will be medicated and unmedicated Schizophrenia/Schizoaffective disorder (SZ/SZA) and community controls.

Study 1:
General Inclusion: 1) age 18 to 65; 2) no substance disorder in past three months (confirmed with saliva drug screens, breathalyzers and saliva swab); and 3) no clinically significant head injury or neurological disease. We will also administer the Wechsler Test of Adult Reading (WTAR) as a measure of premorbid IQ [28]. As in our prior studies, we will exclude participants with scores below 16, which corresponds to an estimated premorbid IQ of less than 70. This will help ensure that participants can understand the basic task instructions. We will ensure that the groups are similar with respect to key demographic variables that could influence task performance (parental SES, age, gender).

SZ/SZA Specific: 1) no medication changes in prior month or anticipated in the next month (or not taking medications, see below), and 2) outpatient or partial hospital. Schizophrenia/Schizoaffective Disorder (SZ/SZA): DSM-5 diagnosis of schizophrenia or schizoaffective.

Unmedicated: Additional inclusion/exclusion criteria for the unmedicated individuals are: 1) not in acute exacerbation that would require immediate hospitalization or treatment; 3) self-reported medication non-compliance for at least the four previous weeks or never having started medications. Unmedicated means antipsychotic-free for schizophrenia and schizoaffective.

Healthy Controls: 1) no DSM-5 Psychosis or Bipolar Disorder; 2) no family history of psychosis or bipolar (we will not exclude for family history of depression (DEP) due to higher base rates).

Study 2:
General Inclusion: 1) age 18 to 65; 2) no substance disorder in past three months (confirmed with saliva drug screens, breathalyzers and saliva swab); and 3) no clinically significant head injury or neurological disease. We will also administer the Wechsler Test of Adult Reading (WTAR) as a measure of premorbid IQ [28]. As in our prior studies, we will exclude participants with a score lower than 16, as this is equivalent to an estimated premorbid IQ of less than 70. This will help to ensure that participants are capable of understanding the basic task instructions. We will ensure that the groups are similar in terms of key demographics that could influence task.

SZ/SZA Specific: 1) no medication changes in prior month or anticipated in the next month (or not taking medications for Study 1, see Medications below), and 2) outpatient or partial hospital. Schizophrenia/Schizoaffective Disorder (SZ/SZA): DSM-5 diagnosis of schizophrenia or schizoaffective.

Healthy Controls: 1) no DSM-5 Psychosis or Bipolar Disorder; 2) no family history of psychosis or bipolar (we will not exclude for family history of DEP due to higher base rates).

Study 3: To obtain an extremely broad sample, we will simply require that online participants are residents of the U.S. between 18-65 years of age, but will collect information on a range of potentially influential characteristics (previous mental and physical illness, head trauma, substance use).

D.2.a. Exclusion Criteria
Participants will be excluded if they do not meet the inclusion criteria outlined above. Participants will also be excluded if they have had and electro-convulsive therapy (ECT) in the past 6 months, transcranial magnetic stimulation (TMS) therapy in past 3 months, transcranial direct current stimulation (tDCS) in the past 3 weeks, or transcranial alternating current stimulation (tACS) in the past 3 weeks.

D.2.b. Ethical Considerations
Washington University's Human Research Protection Office will serve as the Single IRB for the study. The protocol, informed consent document, and supporting information must be approved by the IRB, and the IRB Reliance Agreements deferring oversight to Washington University's IRB must be fully executed, before study initiation at any other site.

D.2.c. Subject Recruitment Plans and Consent Process

Study 1 Pilot Phase Recruitment: Participants will be recruited via MTurk or Prolific, as well as from individuals who participated in previous studies.

Study 1 Pilot Phase Consent :

Mturk/Prolific (online group; for participants completing all procedures remotely): Consent will always be obtained before a person participates. Participants will read an on-line consent form. This consent will also contain a detailed description of all study procedures, as well as any possible risks and/or benefits. Participants will have a choice as to whether or not to proceed to complete computer tasks, and choosing to proceed after reading the consent and explicitly "clicking" acknowledgment of the information in the consent explanation will be taken as providing informed consent. The on-line nature of the study requires that we seek a waiver for written documentation of consent, though the electronic procedures and record will serve as documentation of the individual's choice to proceed with the study following reading the consent information.

In person (in-person group; for participants coming in to the study site to complete the procedures): Consent will always be obtained before a person participates. For all participants, the consent form will be explained to participants as well as having them read the consent themselves. This consent will contain a detailed description of all study procedures, as well as any possible risks and/or benefits. Initial consent will be obtained either by the Study PI or one of the trained research staff. No participants will be under legal commitment at the time of their consent and participation in the study, in accordance with Missouri statutes. Consent will be documented by having the participant and the individuals obtaining consent both sign and date the consent form.

As needed and required by institutions, we will utilize remote consent and assessment sessions to be conducted by phone, Zoom, or another viable electronic means of communication. A consent form will be emailed or sent to subject after phone screening. (see below). This form will be read by the engaged research team member to the participant on the preferred virtual interface. The subject will sign the consent and send either the original or the electronic signed the document to the researcher before procedures begin.
Researchers will notify participants that use of these means of communication introduce privacy risks. This should be communicated prior to starting the remote visit.

Consent waiver for screening interview (in-person group only): We will seek a waiver of the

requirement to obtain written informed consent for the phone screening phase, as we have done successfully in the past for brief telephone screening phases of a project.

Study 1 and Study 1 Behavioral Pilot Recruitment: Individuals with schizophrenia and schizoaffective disorder will be recruited from three sources: 1) those individuals participating in our other R01 projects; 2) individuals who are attending outpatient facilities in the community; or 3) individuals who respond to our outreach. As with the participants with psychiatric disorders, controls will be recruited from among those individuals participating in our other projects, as well as from the general population in the surrounding area through advertisements in local newspapers and TV, and advertisements placed in the same neighborhoods in which the participants with schizophrenia tend to live (e.g., on buses, metros, local shops). Participants will be recruited from outpatient community mental health centers and day treatment programs with whom we have long standing contacts, from psychiatrists treating patients with psychosis and mood pathology at the PIs University and at private offices. We also use a variety of advertising approaches, including: a) Volunteers for Health, a recruitment office supported by the Washington University Medical School that maintains online advertisement systems; b) advertisements in the NAMI newspaper; and c) mailing and flyers to clinics and psychiatrists. The individuals with schizophrenia, schizoaffective disorder, bipolar disorder and depression will be outpatients.
Study 1 and Study 1 Behavioral Pilot Consent: Consent will always be obtained before a person participates. For all participants, the consent form will be explained to participants as well as having them read the consent themselves. This consent will contain a detailed description of all study procedures, as well as any possible risks and/or benefits. Initial consent will be obtained either by the Study PI or one of the trained research staff. No participants will be under legal commitment at the time of their consent and participation in the study, in accordance with Missouri statutes. Consent will be documented by having the participant and the individuals obtaining consent both sign and date the consent form

Electronic Consent (econsent)
This study will utilize the WUSTL Redcap econsent feature to obtain informed consent from study participants at all sites. The Redcap Econsent database will be separate from the study database in order to completely separate from any study data including the participant's study identification number. The Redcap econsent database will only contain the information collected as part of the consent process, with no connection to data collected in the future.

The WUSTL Redcap econsent will be completed as a survey, which will place a timestamp on the participant's signature in the completed form. The econsent will be saved only to the recap database and will be printed (on site only) by study staff and kept with other consent forms on site. The econsent will not be downloaded to any desktop, server or cloud-based location. A record of the signed econsent will be provided to the participant. All electronic communication with participants will be done over a secure server.

As needed and required by institutions, we will utilize remote econsent and assessment sessions to be conducted by phone, Zoom, or another viable electronic means of communication. This form will be read by the engaged research team member to the participant on the preferred virtual interface. The subject will sign the consent and be provided with the completed consent document before procedures begin.
Researchers will notify participants that use of these means of communication introduce privacy risks. This should be communicated prior to starting the remote visit.

Consent waiver for screening interview: We will seek a waiver of the requirement to obtain informed written consent for the phone screening phase, as we have done successfully in the

past for brief telephone screening phases of a project.

Study 2 Recruitment: Individuals with schizophrenia and schizoaffective disorder will be recruited from three sources: 1) those individuals participating in our other Research Project Grant (R01) projects; 2) individuals who are attending outpatient facilities in the community; or 3) individuals who respond to our outreach. As with the participants with psychiatric disorders, controls will be recruited from among those individuals participating in our other projects, as well as from the general population in local communities through advertisements in local newspapers and TV and advertisements placed in the same neighborhoods in which the participants with schizophrenia tend to live (e.g., on buses, metros, local shops). Participants will be recruited from outpatient community mental health centers and day treatment programs with whom we have long standing contacts, from psychiatrists treating patients with psychosis and mood pathology at all recruitment sites. We also use a variety of advertising approaches, including: a) recruitment office supported by the universities: b) advertisements in the NAMI newspaper; and c) mailing and flyers to clinics and psychiatrists. The individuals with schizophrenia, schizoaffective disorder, bipolar disorder and depression will be outpatients.

Study 2 Consent: Consent will always be obtained before a person participates. For all participants, the consent form will be explained to participants as well as having them read the consent themselves. This consent will contain a detailed description of all study procedures, as well as any possible risks and/or benefits. Initial consent will be obtained either by the Study PI or one of the trained research staff. No participants will be under legal commitment at the time of their consent and participation in the study, in accordance with Missouri statutes. Consent will be documented by having the participant and the individuals obtaining consent both sign and date the consent form. We will maintain the option of a hard copy and an econsent process in case the redcap server is down.

Electronic (E Consent)
This study will utilize the WUSTL Redcap econsent feature to obtain informed consent from study participants at all sites. The Redcap Econsent database will be separate from the study database in order to completely separate from any study data including the participant's study identification number. The Redcap econsent database will only contain the information collected as part of the consent process, with no connection to data collected in the future.

The WUSTL Redcap econsent will be completed as a survey, which will place a timestamp on the participant's signature in the completed form. The econsent will be saved only to the recap database and will be printed (on site only) by study staff and kept with other consent forms on site. The econsent will not be downloaded to any desktop, server or cloud-based location. A record of the signed econsent will be provided to the participant. All electronic communication with participants will be done over a secure server.

As needed and required by institutions, we will utilize remote consent processes and clinical interview sessions to be conducted by phone, Zoom, or another viable electronic means of communication. A consent form will be emailed or sent to subject after phone screening. (see below). This form will be read by the engaged research team member to the participant on the preferred virtual interface. The subject will sign the consent and send either the original or the electronic signed the document to the researcher before procedures begin.

Researchers will notify participants that use of these means of communication introduce privacy risks. This should be communicated prior to starting the remote visit.

Consent waiver for screening interview: We will seek a waiver of the requirement to obtain written informed consent for the phone screening phase, as we have done successfully in the past for brief telephone screening phases of a project.

Study 3 Recruitment: Online data collection via Mturk, Prolific or similar online platform

Study 3 Consent: Consent will always be obtained before a person participates. Participants will read an on-line consent form. This consent will also contain a detailed description of all study procedures, as well as any possible risks and/or benefits. Participants will have a choice as to whether or not to proceed to complete self-report questions and computer tasks, and choosing to proceed after reading the consent and explicitly "clicking" acknowledgment of the information in the consent explanation will be taken as providing informed consent. For Study 3, the on-line nature of the study requires that we seek a waiver for written documentation of consent, though the electronic procedures and record will serve as documentation of the individual's choice to proceed with the study following reading the consent information.

D.2.d Randomization Method and Blinding
NOT APPLICABLE

D.2.e. Risks and Benefits

Study 1 Pilot Phase

In-person:
Potential Risks: The potential risks of this project include the following. The clinical interview or self-report instruments to be used for the project could induce boredom and fatigue. However, participants may take a break at any time. There is a theoretical possibility (but extremely small) that a breach in confidentiality could occur (break-in at the Wash U or other research offices), though no identifying information will be collected in this on-line study. There are no known risks associated with any of the behavioral or cognitive testing procedures to be used in the proposed project. There are no alternative procedures or methods available that offer lower risk or cost and provide comparable information to that obtained by the procedures described in this proposal.

Online:
Potential Risks: The potential risk of this project include boredom from task completion. There are no known risks associated with any of the behavioral or cognitive testing procedures to be used in the proposed project. There are no alternative procedures or methods available with less risk or cost that provide comparable information to that obtained by the procedures described in this proposal.

Potential Benefits of the Proposed Research to Human Subjects and Others: There will be no direct benefits to individuals for participating in this proposed project. The risks of this project are relatively small, and not appreciably greater than the risks of ordinary clinical and cognitive assessment.

Study 1 Behavioral Pilot
Potential Risks: The potential risk of this project include boredom from task completion. The self-report instruments to be used for the project could induce boredom and fatigue. There is a theoretical possibility (but extremely small) that a breach in confidentiality could occur (break-in at the Wash U or other research offices).There are no known risks associated with any of the behavioral or cognitive testing procedures to be used in the proposed project. There are no

alternative procedures or methods available with less risk or cost that provide comparable information to that obtained by the procedures described in this proposal.

Potential Benefits of the Proposed Research to Human Subjects and Others: There will be no direct benefits to individuals for participating in this proposed project. The risks of this project are relatively small, and not appreciably greater than the risks of ordinary clinical and cognitive assessment.

Study 1

Potential Risks The potential risks of this project include the following. The clinical interview or self-report instruments to be used for the project could induce boredom and fatigue. However, breaks in interview sessions will be given whenever the subject requests or obvious fatigue is observed. In the EEG/ERP experiments, there is a theoretical risk of disease transmission via the electrodes; this is no greater than the risks involved in a clinical EEG procedure, and there are no documented cases of disease transmission via scalp EEG electrodes in research settings. In addition, some subjects may experience some minor, temporary discomfort during the application of the electrodes. In our experience, approximately 1 out every 500 subjects elects not to participate because of fear of the electrode application procedure or due to claustrophobia in the recording chamber. There is a theoretical possibility (but extremely small) that a breach in confidentiality could occur (break-in at the Wash U or other research offices).

If during the SCID interview, a participant indicates current suicidal or homicidal ideation sufficient to warrant concern, the Interviewer will consult with an investigator at that site who is a clinician. At Washington University, Dr. Barch is a licensed Clinical Psychologist. At the Maryland Psychiatric Research Center, Dr. Gold is a licensed Clinical Psychologist. At the University of California at Irvine, Dr. Carter is a Psychiatrist. At the University of Chicago, Dr. Erickson is a licensed Clinical Psychologist. At the University of Minnesota, Dr. MacDonald is a licensed Clinical Psychologist. When necessary, appropriate referrals to either immediate Emergency Room treatment at the University (each site is within a University with a hospital) or referrals for mental health care in the participant's local community will be arranged. In addition, if a previously undiagnosed or untreated psychiatric, neurological or any other health condition is identified, appropriate referrals for treatment will be provided to the participant.

Econsent: There is a breach of confidentiality and privacy risk in the completion of the econsent process and the virtual /remote session. There are no known risks associated with any of the behavioral or cognitive testing procedures to be used in the proposed project;

We will be recruiting individuals who have taken themselves off medications or who have never started medications. There is the potential risk that such individuals could see our wanting to recruit them as an implicit support or encouragement of their not taking their medications.

There are no alternative procedures or methods available with less risk or cost that provide comparable information to that obtained by the procedures described in this proposal.

Potential Benefits of the Proposed Research to Human Subjects and Others: There will be no direct benefits to individuals for participating in this proposed project. The risks of this project are relatively small, and not appreciably greater than the risks of ordinary clinical and cognitive assessment.

Study 2

Potential Risks: The potential risks of this project include the following. The clinical interview or self-report instruments to be used for the project could induce boredom and fatigue. However, breaks in interview sessions will be given whenever the subject requests or obvious fatigue is observed. There is a theoretical possibility (but extremely small) that a breach in confidentiality could occur (break-in at the Wash U or other research offices). There is a breach of confidentiality and privacy risk in the virtual consent process and virtual session. There are no known risks associated with any of the behavioral or cognitive testing procedures to be used in the proposed project. There are no alternative procedures or methods available with less risk or cost that provide comparable information to that obtained by the procedures described in this proposal.

If during the SCID interview, a participant indicates current suicidal or homicidal ideation sufficient to warrant concern, the Interviewer will consult with an investigator at that site who is a clinician. At Washington University, Dr. Barch is a licensed Clinical Psychologist. At the Maryland Psychiatric Research Center, Dr. Gold is a licensed Clinical Psychologist. At the University of California at Irvine, Dr. Carter is a Psychiatrist. At the University of Chicago, Dr. Erickson is a licensed Clinical Psychologist. At the University of Minnesota, Dr. MacDonald is a licensed Clinical Psychologist. When necessary, appropriate referrals to either immediate Emergency Room treatment at the University (each site is within a University with a hospital) or referrals for mental health care in the participant's local community will be arranged. In addition, if a previously undiagnosed or untreated psychiatric, neurological or any other health condition is identified, appropriate referrals for treatment will be provided to the participant.

Potential Benefits of the Proposed Research to Human Subjects and Others: There will be no direct benefits to individuals for participating in this proposed project. The risks of this project are relatively small, and not appreciably greater than the risks of ordinary clinical and cognitive assessment.

Study 3

Potential Risks: The potential risks of this project include the following. The clinical interview or self-report instruments to be used for the project could induce boredom and fatigue. However, participants can take a break whenever they want. There is a theoretical possibility (but extremely small) that a breach in confidentiality could occur (break-in at the Wash U or other research offices)., though no identifying information will be collected in this on-line study. There are no known risks associated with any of the behavioral or cognitive testing procedures to be used in the proposed project. There are no alternative procedures or methods available with less risk or cost that provide comparable information to that obtained by the procedures described in this proposal.

Potential Benefits of the Proposed Research to Human Subjects and Others: There will be no direct benefits to individuals for participating in this proposed project. The risks of this project are relatively small, and not appreciably greater than the risks of ordinary clinical and cognitive assessment.

E. Study Procedures

E.1. Screening for Eligibility

For Studies 1 pilot and Study 1 Behavioral Pilot phase (in-person group), 1 and 2, initial screening for eligibility will occur through a phone screen, with review of eligibility criteria at the start of the first in-person session.

For Studies 1 pilot phase (in-person group), Study 1 Behavioral Pilot, 1, and 2, after consent, we will also administer the Wechsler Test of Adult Reading (WTAR) as a measure of premorbid IQ [28]. As in our prior studies, we will exclude participants with a score lower than 16, as this is equivalent to an estimated premorbid IQ of less than 70.

For the Study 1 Pilot Phase (online group) and Study 3, initial computer screens for the online Mturk or Prolific or similar platform will assess the age eligibility criteria.

E.2. Schedule of Measurements

Study 1 Pilot Phase (online group): The online pilot study will include various combinations of the questionnaires and the tasks: Motion Perception, Language Perception, Uncertainty Task, Three Choice Prob Reversal, Kamin Blocking, Minesota Trust Game, and Dynamic Trust Game. The questionnaires and tasks are described below. The online participants will complete one to three sessions, which can range in time from 15 minutes to 2 hours.

Study 1 (Pilot Phase in-person group) and Study 1 (Full study): Each participant will complete 4 sessions (2-3 hours each) within 3 weeks (as in our prior studies).

Study 1 Behavioral Pilot: Each participant will be asked to complete 1 to 2 sessions within three weeks. At the UCI site, an additional clinical interview session will be conducted.

Participants will also be screened for drug and alcohol use at each session and will be excluded from further participation in the study if they test positive for either drug or alcohol use. Alcohol use will be tested with the AlcoHawk ABI Digital Breathalyzer, which has disposable mouthpieces. The cut off will be 0.05. Drug use will be tested with Confirm Biosciences Saliva test Confirm - 6 Panel Oral Fluid Test Device F with Saturation Indicator, which will screen for THC, Cocaine, Methamphetamine, Amphetamine and Morphine. Participants will be asked to urinate in a cup and the test card is dipped in the cup. If a participant has a positive test result, they will be asked to provide saliva swab to confirm results in a Drager unit. Participants will be excluded for a positive test for any one of the 5 drugs listed above. Drug and alcohol testing will be the first thing done in each testing session. If a participant tests positive for either alcohol or drugs, they will not be included in the study and will not participate in any further procedures. In virtual session one, we will not administer drug screens.

The first session includes clinical, self-report, and proxy function. Sessions 2-4 are for cognition and EEG measurements. We will administer the maximum # of trials necessary to determine whether tasks can be shortened without losing construct validity and to facilitate EEG analyses. Thus, initial tasks will be longer than ideal, and analyses from all 3 studies will be used to shorten tasks. Tasks will be counterbalanced across session within each group, with 2-3 tasks/session. During the tasks described below, 64-channel EEG will be recorded from DC–400 Hz using a 1000 Hz sampling rate. Offline data preprocessing will be conducted using EEGLAB159 and ERPLAB160, including segmenting, baseline correction, ICA-based artifact correction, artifact rejection, and filtering.

Enrolled patient population: Patients may provide optional informant information for Specific Level of Functioning Scale – Informant and WHODAS Proxy. We ask the patient population for

permission to contact a proxy (i.e. social worker, case manager, or family member) to complete questionnaires about the patient.

Tasks for Study 1 (In-Person Pilot and Full Study):

Sine Wave Speech: Sine wave speech (SWS) stimuli are made by replacing the formants (main bands of energy) in speech with pure tone whistles. Participants will listen to intelligible and unintelligible SWS and indicate whether they heard speech or not, along with a confidence judgement. After Run 1, participants may be asked if they noticed any words in the distractor stimuli. We will test their ability to discriminate between intelligible SWS and unintelligible SWS (d'), their bias in classifying speech, and accuracy (number of key words correct). Participants are then explicitly told that there is actual speech in some of the stimuli and they will be exposed to some pre-degradation speech templates, and the task will be repeated (Run 2). We will again test their ability to discriminate between intelligible SWS and unintelligible SWS (d'). This task takes 30-40 minutes to complete.

Motion Change Task: This task uses a random dot kinematogram paradigm in which dots moving coherently in a single direction are mixed with randomly moving dots. On 50% of trials, the direction of coherent motion changes by 90° midway through the trial. Participants are asked to report the direction perceived at the end of the trial. This allows us to measure how much an existing representation can be updated based on new information. There are between 50 and 150 trials, and this task takes approximately 10 minutes.

Belief Updating Task: Participants are presented with 3 decks of cards on a computer screen and told that each deck includes both winning (1100 points) and losing (250 points) cards, but that some decks win more often than others. They are instructed to find the best deck (i.e., the one with the highest probability for reward) and that the best deck might change, whereupon they should try and find the new best deck. Participants' overarching goal is to win the most points they can. To enhance engagement and motivation, they are offered a $2 incentive for winning more points than 75% of other participants. There are approximately 120-160 trials divided into blocks (typically 4), with an optional break in between each block. Before beginning the main task, participants complete practice trials. The task was designed to elicit the experience of volatility in several ways. First, the best deck changes when a participant selects it in 9 of 10 consecutive trials (reversal events). Second, the decks yield 90%, 50%, and 10% reward for blocks 1 and 2 (~60-80 trials) and then these contingencies change at the start of block 3 to 80%, 40%, and 20% (contingency transition). The goal of this task structure was to make it more difficult for participants to discern whether a loss was due to probabilistic noise or due to the best deck changing. By beginning the task with a 90-50-10 probability structure, participants are expected to develop stronger beliefs about the task environment, making the contingency transition more confounding, thereby increasing the volatility experience. Participants can earn up to a $2 incentive for the task.

Uncertainty in Learning Task: Participants are instructed to place a bucket at a horizontal position on the ground to maximize the chances of catching bags that would be dropped from a helicopter. On each trial, a bag is dropped from the helicopter at the top of the screen with a probabilistic horizontal displacement. During training, the helicopter is visible to participants and they can observe the variability from trial to trial in the bags around it. Bucket placement is accomplished using a gamepad that contains right and left buttons that control the movement of the bucket from its previous position and a separate button to confirm bucket placement. During the subsequent test phase, the helicopter is obscured by clouds, forcing participants to infer its position based on previously observed bag locations. To minimize the need for working memory necessary to represent the previous bag location the task provided a visual depiction of the

prediction error from the previous trial, spanning the range from the middle of the previous bucket position to the most recent bag location. . This task uses 100-200 trials and takes between 20 and 25 minutes to complete.

Kamin Blocking (Conceptual Beliefs, no (in)Stability): Participants are asked to imagine that they are allergists required to learn the causes of allergic reactions in a fictitious patient[29]. They initially learn that one specific food is consistently followed by an allergic response (e.g., a banana). Then they are shown that food and another food together (e.g., a banana and a mushroom) and are told that an allergic response occurred. HC individuals fail to learn that the second food (the mushroom) predicts the allergic response: the banana allows them to correctly predict the allergy when the banana and mushroom are presented together, so no PE is generated and no learning about the mushroom occurs. On later trials, when participants are shown the mushroom alone and then told that an allergic response occurred, a PE brain response is observed and they learn that the mushroom predicts an allergic response[29].

Minnesota Trust Game: This paradigm is designed to distinguish between sources of distrust such as rational mistrust and spite sensitivity (i.e., a fear of spite from a partner). It does this by examining mistrust in choices when a partner, or second mover, is disincentivized to betray the participant, or first mover. First movers decide to take a small, safe amount of money for both players ($10, for example) or let the second mover decide between two options. The second mover can choose between a mutual option with an equal, larger amount of money for both ($20, for example) or an uneven amount where the second mover gets some temptation (T) and the first mover receives some adverse payoff (AD). There are two conditions based on the temptation: when T=$25 (for example), the second mover is incentivized to betray the first mover to get a larger reward, which we refer to as Rational Mistrust; when T=$15 (for example), the second mover is disincentivized to betray the player, which we refer to as Suspiciousness. Adverse payoffs (AD) vary across trials -- from losing money to winning more than $10 to identify a boundary between trusting and not trusting. Participants played as the second mover first and then reversed roles. Participants are paid based on their choices from a random trial in both games, with the value of $1 being specified before the experiment, and which is matched with the choice of a previous participant in the study. No deception is required to conduct this experiment. Participants earn between $0 and $1 on this game. This task uses 75 to 150 trials and takes approximately 20 minutes to complete

Dynamic Trust Game: Participants complete trials of the Trust Game with three presumed online partners. Unbeknownst to participants, these partners are preprogrammed, slowly drifting in their reward rate throughout the task, requiring participants to adjust their choice policy to optimize reward continually, the amount of money a partner reciprocates back to the participant. At the start of each trial, participants are paired with one of three online partners, and can invest $1 or $10. The invested money is then quadrupled in value, and the partner receives the quadrupled sum. Partners then decide to return anywhere from 0-50% of the investment such that participants can lose all of the initial investment (0% return), double their investment (50% return), or receive any outcome in between. Social partners gradually reverse their payouts, requiring participants to adjust their choice policy continually. Trials commence with a partner pairing phase in which the computer selected a partner for the current round. Participants are then given up to 4 seconds to indicate their investment and then make a prediction about how much money they believed their partner would return, before observing trial outcomes. Participants are told that they will receive the outcome of 1 randomly chosen trial, which will be their bonus payment in addition to whatever the base compensation is for completing the study. The minimum is $0 and the max is $20, but most people's bonus payment was between $7-$15. This task is run with 150 to 250 trials and takes approximately 20 minutes. Participants are debriefed after.

Clinical Assessments for Study 1
Structured Clinical Clinical Interview for DSM-5 (SCID-5) (diagnosis), Psychosis, Mood and Substance Use Modules (SCID)
Brief Psychiatric Rating Scale (BPRS)
Chicago Hallucinations Assessment Tool (CHAT)
Psychotic Symptoms Rating Scales (PSYRATS)
Clinical Assessment Interview for Negative Symptoms (CAINS)

Patient Population Only: There is an optional video recording of the clinical interview, which includes the following scales: SCID, BPRS, CHAT, PSYRATS, CAINS. The video will be stored on WUSTL Box and used across study sites to calibrate ratings and for training purposes of personnel.

Questionnaires for Study 1
Beliefs about Voices Scale
Yale Control over Perceptual Experiences Cale
Peters Delusion Inventory
Green Paranoia Scale
Patient Health Questionnaire (PHQ-9)
Generalized Anxiety Disorder Scale (GAD-7)
Motivation and Pleasure Self-Report Scale (MAP-SR)
Aberrant Salience Inventory
Cannabis Use Disorder Identification/Cannabis Use Problems Identification CUDIT/CUPIT
Alcohol use (AUDIT)
Cannabis Use
WHODAS 2.0 (function) – Self
Current Medications Form
Specific Levels of Functioning Scale (SLOF)– Patient
DSM-5 Symptom Measure

Proxy Questionnaires for Study 1
WHODAS 2.0 Proxy
Specific Levels of Functioning Scale – Informant

Study 2. Study 2 will measure psychometric properties (test re-test reliability, lack of floor, ceiling and practice effects) of computational parameters (Aim 2) on tasks that were optimized in Study 1. At baseline, we will administer tasks optimized in Aim 1 (without EEG), to a new sample of individuals with SZ and Community Controls. Participants will complete two parts, test-retest. The study will comprise 5-6 sessions. Each session is 2.5-3 hours long. Part one comprises 3 sessions: a one-session clinical interview and two task sessions, conducted over an estimated 3-week period. Part two comprises two to three sessions and includes a retest of the tasks in sessions two and three, as well as an additional potential clinical session. All participants will be retested ~28 days after the initial assessment (along with clinical assessment). We will use medicated patients for the retest study as patient stability is critical. Although our goal is to shorten task length using the Study 1 analyses, we still administer longer tasks than optimal for public dissemination in Study 2 so that we can determine (via Monte Carlo simulations) the minimum number of trials required for acceptable reliability. Study 2 will use the same tasks, clinical assessment and questionnaire measures as Study 1, except we will exclude and EEG.

Participants will also be screened for drug and alcohol use at each session and will be excluded from further participation in the study if they test positive for either drug or alcohol use. Alcohol use will be tested with the AlcoHawk ABI Digital Breathalyzer, which has disposable mouthpieces. The cut off will be 0.05. Drug use will be tested with a Confirm Biosciences Saliva test Confirm - 6 Panel Oral Fluid Test Device F with Saturation Indicator, which will screen for THC, Cocaine, Methamphetamine, Amphetamine and Morphine. Participants will be asked to urinate in a cup and the test card is dipped in the cup. Participants will be excluded for a positive test for any one of the 5 drugs listed above. Participants who test positive for THC may be asked for a saliva swab to confirm test result in Drager unit. Drug and alcohol testing will be the first thing done in each testing session. If a participant tests positive for either alcohol or drugs, they will not be included in the study and will not participate in any further procedures. In virtual session one, we will not administer drug screens.

Patient Population Only: There is an optional video recording of the clinical interview which includes the following scales: SCID, MINI, BPRS, YMRS, CAINS, and BDRS. The video will be stored on WUSTL Box and used across study sites to calibrate ratings and for training purposes of personnel.

Patient Population Only: There is an optional video recording of the clinical interview, which includes the following scales: SCID, BPRS,CHAT, PSYRATS, CAINS. The video will be stored on WUSTL Box and used across study sites to calibrate ratings and for training purposes of personnel.

Study 3. In Study 3, all tasks and models from Aim 1 will be evaluated in a large community sample of online participants. Note, Study 3's use of an online platform, such as Mturk or Prolific, is distinct from the Online Optimization Phase. Study 2 will use the same tasks and questionnaire measures as Study 1, but will not include the clinical assessments. Study 3 will also include the following scale: Infrequency Scale, the Intersectional Discrimination Index (IDI) plus supplemental IDI questions, the Macarthur scale of social status, a shortened medicine questionnaire, and a zip code questionnaire. Participants will complete half the task set in session one and the remaining tasks in session two. The order of the tasks will be randomized.

E.3. Safety and Adverse Events

Safety and Compliance Monitoring

We will utilize a Data and Safety Monitoring Plan. The PIs will monitor the study for any adverse events, although we do not anticipate any given the nature of the study.

Safety reviews will be performed on an annual basis, or as needed, by the Principal Investigators and engaged members of the research team.

All serious adverse events will be reported to the IRB within the reporting requirement timelines. Should there be a serious adverse event that occurs that increases the risks to the participants, the study will be stopped, an investigation will be conducted, and a findings report will be generated before the study is resumed.

F. Statistical Plan

F.1.Sample Size Determination and Power

As shown in our pilot data, we saw both medium and large effect sizes. However, it is well known that pilot data with smaller sample sizes can overestimate effect size, and thus it is important not to assume that we will have large effects. Thus, our sample size was selected to provide 80% or

greater power to detect medium effect sizes, assuming the need for multiple comparison correction that reduces alpha from .05 to as low as .01. For both Study 1 and Study 2 (N=325), these samples provide power of greater than .80 (a <.01) to detect at least moderate effect sizes (e.g. d≥.39, f≥.21) in hypothesized group differences across tasks and relevant group by condition interactions, as well as correlations of computational and non-computational task performance metrics with functioning and symptoms in the psychosis group (N=200, r≥.24) in. It is crucial to have this level of power to explicitly assess whether model-based parameters are more precise and sensitive than standard behavioral methods across a range of common mental disorders, and to have the capacity to predict clinical symptoms and real-world functioning. The most challenging discrimination -- establishing a significant difference between a correlation of a symptom measure and computationally derived metric versus a standard task metric among psychosis patients (N=200) -- is powered at .80 (a<.01) to detect differences in r of about .20 at low correlations between the standard and computational metrics (*r*=.3), and .18 at moderate correlations (*r*=.5). We have similar power at alpha <.05 within just the psychosis group (N=200). Importantly, Study 3 (total N=5000) with power analyses based on the hypothesis generating first N=2000 will have power to detect R^2 increases as small as .01 over-and-above the demographic variables in regression models such as LASSO in those with complete data (N=1480), with greater power in the confirmatory sample (N=3,000; N=2,200 with complete data).

F.2. Analysis Plan

Study 1 Behavioral Data Analysis (Aim 1): Aim 1 tests the hypothesis that model-based parameters for the measurement of belief are more precise and sensitive than standard behavioral methods in assessing deficits across psychotic disorders and determines which measures have the greatest sensitivity and specificity for the severity of hallucinations vs. delusions, consistent with the data presented in section 3.2. Table 1 outlines the specific standard behavioral metrics versus computational metrics that we will compare for each task. We will use nonparametric resampling statistics, including bootstrapping ([e.g., 30, 31, 32]) and randomization tests ([e.g., 33]). Empirical sampling distributions are used for null hypothesis testing and establishing confidence intervals. We will conclude that a parameter estimate (e.g., effect size, beta, correlation) is significantly different from another if it falls outside the 95% CI of the comparison parameter estimate. We will use FDR multiple comparison correction across related contrasts in these analyses and those below. We will also compute the corresponding Bayes factors, especially helpful in quantify the strength of evidence favoring the null hypothesis[34-37]. We will start by comparing the groups (SZ, HC) on standard metrics from each task using a repeated measures (rm) ANOVA with group as a between subject factor, different conditions as repeated measures, and accuracy or reaction time (RT) as dependent measures. We illustrate this approach using the Three Choice Probabilistic Reversal Task: For this task, we will examine accuracy (likelihood of picking rewarded outcome), as a function of block, reversal phase, and reward probability. Next, we will compare the groups on estimated parameters from the Hierarchical Gaussian Filter (HGF) model, using the computational parameters as the dependent variable (e.g., volatility prior [m_3], flexibility of belief updating [ω_3]. We predict that effect sizes in the rmANOVAs will be greater for the computationally derived parameters than for standard metrics. For example, we predict that SZ will show reduced accuracy compared to HC on the task, particularly after reversals. However, we also predict that patients will show higher volatility priors and reduced belief updating compared to HC, with larger effect sizes. We will use the empirical sampling distributions described above to generate 95% Cis around the effect sizes, and would conclude that the effect sizes are different if the accuracy effect size falls outside (on the smaller side) the 95% Cis for the group difference effect sizes for volatility priors and/or belief updating computation metrics. We can further examine the specificity of these effects by using regressions, asking, for example, whether SZ continue to be impaired compared to HC on WM

once attention or general cognitive ability are taken into account.

To test hypotheses about *relationships to hallucinations and delusions (and other symptoms for specificity)*, sex as a biological variable, and other demographics, we will use hierarchical regressions to determine the ability of a standard metric versus a computational metric to predict symptom domains across or within groups. Step 1 uses the metric/parameter to predict the symptom, as well as dummy codes for diagnostic group and demographic comparisons. Step 2 adds interaction terms between metric/parameter and the group dummy codes. If these interaction betas account for additional variance, it indicates a different relationship between the metric/parameter and symptom across groups. We will use bootstrapping to determine if regression weights vary significantly as a function of metric/parameter. As one example, we would predict that delusions should be related to performance on the reversal learning task. We would compare a regression using accuracy after reversal versus the belief updating metric [ω_3]. We predict that the belief updating metric will be more strongly related to delusions (either clinician-rated or self-reported) than the standard behavioral metric. We will again use the empirical sample distributions to establish 95% confidence interval (CI), and will conclude that the association with belief updating versus accuracy are significantly different if one falls outside the 95% CI of the comparison parameter estimate. Interactions with diagnostic groups will allow us to determine whether such relationships vary by group. To address the key question of whether one type of symptom (hallucination vs. delusion) is more related to performance on a certain type of task, we will conclude that a parameter estimate (e.g., beta, correlation) is significantly different from another if it falls outside the 95% CI of the comparison parameter estimate. We can compare the association of hallucinations vs. delusions to a particular metric (e.g., belief updating metric [ω_3] or initial direction of motion bias from the motion task), as relevant to key questions #2 and #3. To determine if a particular symptom domain is more related to performance on one type of task (e.g., social vs. non-social beliefs), we will use the same approach. For example, we can examine whether delusions are more associated with belief updating on the Dynamic Trust Game (social beliefs) versus the Probabilistic Reversal Task (nonsocial beliefs), as relevant to key questions #1, #4, and #5 from.

Hierarchical Modeling: To better understand how computational constructs translate across tasks we will construct a multi-task Bayesian hierarchical model composed of best-fitting models from each task. This joint modeling approach[38] allows the hierarchical model to estimate the posterior covariance capturing relationships between parameters across tasks, or to include latent factors that themselves jointly influence multi-task parameters. We will fit the model in Hierarchical Sequential Sampling Modeling (HSSM) (our Python-based toolbox) using Markov Chain Monte Carlo (MCMC) with the No-UTurn-Sampler to half of the dataset and estimate covariance parameters that provide the best cross-task computational mappings, and then validate the fit of the model in the held-out half of the data. For example, such an analysis may indicate shared relationships between volatility/belief updating parameters across reversal learning, cannon, and dynamic trust tasks, which would be indicative of a task general belief updating construct. Beyond specific shared computations across tasks, the model will be used to evaluate the degree to which behavioral variability in task performance measured in individual tasks can be explained by variability in the others. Finally, the hierarchical model will be extended to include symptoms, allowing us to evaluate how multi-task computational constructs relate to mental health constructs. One potential pitfall, given the breadth and depth of our dataset, is that we may experience convergence issues when fitting a single model to all task data using MCMC. We have accrued experience in handling such situations with alternative approximate approaches that we would deploy: for example, in HSSM we could use variational inference instead of MCMC, and/or we can estimate parameter covariance for pairs of tasks, rather than for the entire battery at once.

Study 1 Behavioral ERP Analysis (Aim 4): Hierarchical regressions will test specific hypotheses about ERP measures in relation to computational parameters. Step 1 will use the parameter to predict the ERP measure, with dummy codes for diagnostic group. Step 2 adds interaction terms between parameter and group dummy codes. We will compare each of these hypothesized relationships to standard behavioral metrics in the same task (Table 1), with the prediction that computational metrics will be more strongly associated to the EEG/ERP measure than standard metrics. This will be assessed using the same approach described above (empirical sampling distributions to generate 95% Cis for each parameter estimate.

Data Analysis for Aim 2: We will use resampling methods where appropriate to estimate psychometric indices. We will also compare conventional single-participant parameter estimates with hierarchical Bayesian estimates, which can yield substantially greater validity and reliability[39-41]. Ceiling/Floor Effects: Reliability of individual difference measures is adversely affected by ceiling and floor effects, which also undermine measuring treatment effects. For each of the standard and computational metrics, we will examine % of participants scoring within 1 standard deviation (SD) of the highest or lowest possible value. We do not anticipate this being a problem, as our Online Optimization Phase will explicitly focus on ensuring a wide range of performance values. Test-retest: We will use generalizability theory to compute ICCs for standard and computational metrics, allowing us to take into account symptom levels across testing sessions, as well as demographic and biological characteristics such as sex. We will conclude that the ICCs for a standard metric versus a computational metric are different if one falls outside the 95% CI of the other. We can also compare these values across metrics/parameters with different numbers of trials to assess how much tasks can be shortened. Practice Effects: We will examine practice effects by computing change scores across the two sessions. Resampling procedures will provide confidence intervals and allow us to compare the magnitude of practice effects across standard metrics versus computational parameters and for differing numbers of trials.

Data Analysis for Aim 3: Analyses will occur in three stages. **Stage 1** (n=100) will focus on the quality of the data being collected and provide early feedback as to the robustness of our questionnaires and tasks and recruitment methods. **Stage 2** (n=1,900) will confirm Stage 1 data quality analyses, and examine sample representativeness and parameter estimation from online data. Throughout we will monitor demographics patterns to adhere to U.S. Census estimates. We will combine Stage 1 and 2 to examine whether the tasks are too long, such that meta-learning or fatigue effects are drawing parameter estimates down over time, rather than increasing reliability. We will accomplish this by comparing sliding windows of trials (30% at a time, advancing in 5% intervals) to evaluate secular trends. We will also examine convergence of parameter estimates (e.g. precision) across task domains. We will then use Group LASSO (*glasso*) penalized regression[42] with 10-fold cross-validation to determine the relationships between self-reported experiences of the positive dimension of psychosis (viz. SPQ Cognitive-Perceptual Scale, Beliefs about Voices Scale[43, 44], Yale Control Over Perceptual Experiences (COPE) scale[45], Peters Delusion Inventory[46, 47] and the Green Paranoia Scale[48]) and our behavioral metrics above demographic variables including sex. We will test specificity by examining other dimensions of psychopathology. Another critical test will be the extent to which computational parameters as compared to standard performance metrics (Table 1) are correlated with symptom-relevant measures. These observations will be pre-registered with the Open Science Foundation. **Stage 3** (n=3,000) will echo Stage 2 in terms of recruitment and substantive analyses, and test the replicability of the pre-registered point estimates. We are adequately powered to use conservative corrections for multiple comparisons, such as exploratory analyses examining other dimensions of psychopathology, and to again utilize both within and out-of-

sample generalization procedures.

F.3. Statistical Methods
See Above

F.4 Missing Outcome Data
Individuals with missing outcome data will not be able to be included in analyses, unless we use imputation methods.

G. Data Handling and Record Keeping

G.2 Confidentiality and Security
All data will be stored under code numbers without participant identifiers. All diagnostic interview, self-report questionnaires, and behavioral testing are conducted in private, away from other individuals. All staff members are required to sign an annual letter indicating that they understand the requirement to maintain absolute confidentiality, and that they will be terminated immediately if a breach of confidentiality occurs. Care will be taken to recode data necessary to ensure they are not individually identifying before public release. All electronic data will be stored password encrypted files on computer systems (e.g., REDCAP, BOX) that require login authentication.

G.2 Training
All staff members are required to participant in CITI and HIPPA training. All staff members are required to sign an annual letter indicating that they understand the requirement to maintain absolute confidentiality, and that they will be terminated immediately if a breach of confidentiality occurs.

G.3 Case Report Forms and Source Documents
There will be no case report forms. All data will be collected electronically using REDCAP or PsychoPy.

G.4Records Retention
Data will be retained for the required period.

G.5 Performance Monitoring
Site performance in recruitment will be monitored through weekly reports generated through REDCAP, weekly PI calls that review performance and weekly or bi-weekly coordinator calls.

H. Study Administration

H.1 Organization and Participating Centers
There are 8 participating sites: Washington University in St. Louis, Brown University, University of California, Irvine, University of California, Davis, University of Chicago, Yale University, Maryland Psychiatric Research Center, and University of Minnesota. Washington University is the coordinating site. The PIs have weekly phone calls, as do the coordinators (lead by the Washington University coordinator). Brown University will primarily be involved in piloting and analysis. Yale and University of California at Davis will primarily be involved in analysis. Data collection for Studies 1 and 2 will occur over the other 5 sites. University of Minnesota will manage Study 3.

H.2 Funding Source and Conflicts of Interest

The study will be funded by the National Institute of Mental Health.

H.3. Committees

Not applicable

H.4 Subject Stipends or Payments

Study 1 Pilot: 4 sessions, each ~2.5 hours, $225 (In addition, participants are able to earn small amounts of money on two of the effort/reward tasks: $1 to $10 for one task; $1 to $15 for other task), up to a total of $2 to $25 across the two tasks.) Up to $250 total payment.

Study 1 Pilot/online: $6.50 per hour, pay will be prorated at the rate of $6.50, with a minimum payment of $1. Length of sessions will vary across sites. If they do not complete the study, they will be pro-rated at the rate of $6.50 per hour, with a minimum of $1

Study 1: 4 sessions, each ~2.5 hours, $225 (In addition, participants are able to earn small amounts of money on two of the effort/reward tasks: $1 to $10 for one task; $1 to $15 for other task), up to a total of $2 to $25 across the two tasks.) Up to $250 total payment.
Study 1 Behavioral Pilot: 1-2 sessions, approximately 2 hours each. Participants will be paid according to site standards (between $10-$15 per hour) with the opportunity to earn up to $23 in bonus money during three of the tasks.

Study 2: 8 sessions (4 at baseline, 4 at 1 month follow up), each ~2.5 hours, $450. UCI site will add an additional consenting session, which will pay $25 for participation. In addition, participants are able to earn small amounts of money on two of the effort/reward tasks: $1 to $10 for one task; $1 to $15 for other task), up to a total of $2 to $25 across the two tasks. Up to $500 total payment.

Study 3: Participants will be paid through an online platform, such as Mturk or Prolific. The duration of the study is approximately 2 hours in completion, estimated 1 hour per session. Participants may have the opportunity to earn an additional bonus payment on certain tasks. Payment rates to participants will be determined by the site.

Informant Reports: $20 per person

H.5 Study Timetable

We anticipate that Study 1 will take ~20 months (data acquisition & analysis) after a 6 month pilot phase and Study 2 will take ~28 months. The slightly longer time for Study 2 reflects the re-test component. Study 3 will run in parallel with Studies 1 & 2, during years 2-4.

I. Publication Plan

The results of the study will be published as they become available after completion of each of the three proposed studies.

J. References

1. Yates, K., et al. JAMA Psychiatry, 2019. *76*(2): 180-189.
2. Coid, J.W., et al. JAMA Psychiatry, 2013. *70*(5): 465-71.
3. Kadakia, A., et al. J Clin Psychiatry, 2022. *83*(6).

4. Rossi-Goldthorpe, R.A., et al. PLoS Comput Biol, 2021. *17*(10): e1009453.
5. Jeganathan, J., et al. Lancet Psychiatry, 2021. *8*(8): 732-738.
6. Liddle, P.F., et al. Front Hum Neurosci, 2022. *16*: 818711.
7. Liddle, P.F., et al. Biol Psychiatry, 2024.
8. Stuke, H., et al. Front Psychol, 2021. *12*: 583637.
9. Feyaerts, J., et al. Lancet Psychiatry, 2024. *11*(7): 497-498.
10. Harding, J.N., et al. Lancet Psychiatry, 2024. *11*(7): 498.
11. Harding, J.N., et al. Lancet Psychiatry, 2024. *11*(4): 295-302.
12. Sheffield, J.M., et al. Biol Psychiatry Glob Open Sci, 2024. *4*(4): 100333.
13. Schmack, K., et al. JAMA Psychiatry, 2022. *79*(1): 75-76.
14. Corlett, P.R., et al. Trends Cogn Sci, 2019. *23*(2): 114-127.
15. Kazinka, R., et al. Biol Psychiatry Cogn Neurosci Neuroimaging, 2024. *9*(4): 429-436.
16. Kazinka, R., et al. Comput Psychiatr, 2022. *6*(1): 60-78.
17. Nassar, M.R., et al. Brain, 2021. *144*(3): 1013-1029.
18. Rossi-Goldthorpe, R., et al. Brain, 2024. *147*(8): 2854-2866.
19. Sheffield, J.M., et al. Biol Psychiatry Cogn Neurosci Neuroimaging, 2022. *7*(11): 1149-1157.
20. Suthaharan, P., et al. Sci Rep, 2023. *13*(1): 6084.
21. Suthaharan, P., et al. Res Sq, 2021.
22. Wisner, K.M., et al. Hum Brain Mapp, 2021. *42*(16): 5244-5263.
23. Haarsma, J., et al. J Neurosci, 2023. *43*(47): 7946-7957.
24. Sterzer, P., et al. Biol Psychiatry, 2018. *84*(9): 634-643.
25. Fleming, L.M., et al. Sci Rep, 2023. *13*(1): 4841.
26. Lemonde, A.C., et al. Schizophr Bull, 2023. *49*(4): 1032-1041.
27. Bjarke, J., et al. Nord J Psychiatry, 2020. *74*(7): 497-504.
28. Wechsler, D., *Wechsler test of adult reading*, 2001, The Psychological Corporation: San Antonia, TX.
29. Corlett, P.R., et al. Neuropsychologia, 2012. *50*(14): 3612-20.
30. Chernick, M.R., *Bootstrap methods: A guide for practitioners and researchers (2nd Ed)*2007, New York: Wiley.
31. Mooney, C.Z., et al., *Bootstrapping: A nonparametric approach to statistical inference*. Quantitative Applications in the Social Sciences, ed. M.S. Lewis-Beck. Vol. 95. 1993, Newbury Park: Sage Publications.
32. Efron, B.D., et al., *An introduction to the bootstrap*1994, Norwell, MA: Kluwer Academic Publishers.
33. Edgington, E., et al., *Randomization tests (4th Edition)*2007, Norwell, MA: Kluwer Academic Publishers.
34. Province, J.M., et al. Proc Natl Acad Sci U S A, 2012. *109*(36): 14357-62.
35. Rouder, J.N., et al. Psychon Bull Rev, 2009. *16*(2): 225-37.
36. Wagenmakers, E.J., et al. Psychon Bull Rev, 2017.
37. Wagenmakers, E.J., et al. Psychon Bull Rev, 2017.
38. Kvam, P.D., et al. Psychol Methods, 2021. *26*(1): 18-37.
39. Brown, V.M., et al. Biol Psychiatry Cogn Neurosci Neuroimaging, 2020. *5*(6): 601-609.
40. Ahn, W.Y., et al. J Neurosci Psychol Econ, 2011. *4*(2): 95-110.
41. Wetzels, R., et al. Journal of Mathematical Psychology, 2010. *54*: 14-27.
42. Yang, Y., et al. Statistics and Computing, 2015. *25*: 1129-1141.
43. Chadwick, P., et al. Br J Psychiatry, 1995. *166*(6): 773-6.
44. Chandwick, P., et al. Br J Psychiatry, 2000. *177*: 229-32.
45. Mourgues, C., et al. Schizophr Bull, 2022. *48*(3): 673-683.
46. Peters, E., et al. Schizophr Bull, 2004. *30*(4): 1005-22.
47. Peters, E.R., et al. Schizophr Bull, 1999. *25*(3): 553-76.

48. Green, C.E., et al. Psychol Med, 2008. *38*(1): 101-11.