Behaviour Research and Therapy 101 (2018) 46–57

Contents lists available at ScienceDirect

Behaviour Research and Therapy

journal homepage: www.elsevier.com/locate/brat

Applying novel technologies and methods to inform the ontology of self-regulation

Ian W. Eisenberg[a,*], Patrick G. Bissett[a], Jessica R. Canning[d], Jesse Dallery[f], A. Zeynep Enkavi[a], Susan Whitfield-Gabrieli[j], Oscar Gonzalez[d], Alan I. Green[b], Mary Ann Greene[c], Michaela Kiernan[e], Sunny Jung Kim[b,c], Jamie Li[a], Michael R. Lowe[k], Gina L. Mazza[d], Stephen A. Metcalf[c], Lisa Onken[i], Sadev S. Parikh[a], Ellen Peters[g], Judith J. Prochaska[e], Emily A. Scherer[b], Luke E. Stoeckel[h], Matthew J. Valente[d], Jialing Wu[c], Haiyi Xie[c], David P. MacKinnon[d], Lisa A. Marsch[b,c], Russell A. Poldrack[a]

[a] *Department of Psychology, Stanford University, Stanford, CA 94305, USA*
[b] *Department of Psychiatry, Geisel School of Medicine at Dartmouth, Dartmouth College, Lebanon, NH 03766, USA*
[c] *Department of Biomedical Data Science, Geisel School of Medicine at Dartmouth, Dartmouth College, Lebanon, NH 03766, USA*
[d] *Department of Psychology, Arizona State University, Tempe, AZ 85281, USA*
[e] *Department of Medicine, Stanford University, Stanford, CA 94305, USA*
[f] *Department of Psychology, University of Florida, Gainesville, FL 32611, USA*
[g] *Department of Psychology, The Ohio State University, Columbus, OH 43206, USA*
[h] *National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, MD 20892, USA*
[i] *National Institute on Aging, National Institutes of Health, Bethesda, MD 20892, USA*
[j] *Brain and Cognitive Sciences Department, The McGovern Institute for Brain Research, MIT, Cambridge, MA 02139, USA*
[k] *Department of Psychology, Drexel University, Philadelphia, PA, USA*

ARTICLE INFO

Keywords:
Self-regulation
Ontology
Neuroimaging
Intervention
Obesity
Smoking

ABSTRACT

Self-regulation is a broad construct representing the general ability to recruit cognitive, motivational and emotional resources to achieve long-term goals. This construct has been implicated in a host of health-risk behaviors, and is a promising target for fostering beneficial behavior change. Despite its clear importance, the behavioral, psychological and neural components of self-regulation remain poorly understood, which contributes to theoretical inconsistencies and hinders maximally effective intervention development. We outline a research program that seeks to define a neuropsychological ontology of self-regulation, articulating the cognitive components that compose self-regulation, their relationships, and their associated measurements. The ontology will be informed by two large-scale approaches to assessing individual differences: first purely behaviorally using data collected via Amazon's Mechanical Turk, then coupled with neuroimaging data collected from a separate population. To validate the ontology and demonstrate its utility, we will then use it to contextualize health risk behaviors in two exemplar behavioral groups: overweight/obese adults who binge eat and smokers. After identifying ontological targets that precipitate maladaptive behavior, we will craft interventions that engage these targets. If successful, this work will provide a structured, holistic account of self-regulation in the form of an explicit ontology, which will better clarify the pattern of deficits related to maladaptive health behavior, and provide direction for more effective behavior change interventions.

1. Introduction

Self-regulation is a broad construct that refers to the ability to modulate behavior in service of long-term goals. When we sleep through our alarm repeatedly, we blame our poor self-control; when others exercise regularly and maintain their diet, we are impressed by their restraint and commitment. Across many different subdomains of psychology, self-regulatory ability is described as an important dimension of individual variability and is often invoked to explain a wide variety of maladaptive or suboptimal behaviors. While the field has

* Corresponding author.
E-mail address: ieisenbe@stanford.edu (I.W. Eisenberg).

https://doi.org/10.1016/j.brat.2017.09.014
Received 8 April 2017; Received in revised form 25 September 2017; Accepted 30 September 2017
Available online 05 October 2017
0005-7967/

created a number of different metrics to summarize self-regulation (e.g., delay of gratification: Mischel, Shoda, & Rodriguez, 1989; grit: Duckworth & Quinn, 2009; impulsivity: Citation, Whiteside, & Lynam, 2001; Stanford et al., 2009), the underlying claims are consistent—-people differ in some ability or abilities in a way that affects how successfully they can modulate their behavior to achieve their long-term goals.

Numerous studies have suggested that differences in self-regulatory ability are related to differences in health, education, social networks, and economic well-being (Mischel et al., 1989; Moffitt et al., 2011). Early exposures in utero and in adolescence to substances, such as nicotine, are also associated with adverse behavioral effects on cognitive functions associated with self-regulation (England et al., 2017). Self-regulation is related to health risk behaviors, including poor diet, physical inactivity, tobacco and other substance use (Bickel, Odum, & Madden, 1999; Epstein, Salvy, Carr, Dearing, & Bickel, 2010; de Ridder & de Wit, 2006)—behaviors that account for as much as 40% of the illness, suffering, and early death related to chronic disease.

Given the high societal cost of poor health linked with health risk behavior, an array of interventions have been developed to promote the initiation and maintenance of health behavior change (Krebs, Prochaska, & Rossi, 2010). However, the development of these interventions has been largely siloed (e.g. focusing on one disease at a time) and is typically intended to engage multiple mechanisms of behavioral change, without systematic examination of how the interventions actually work. An alternative "experimental medicine" approach, adopted in the Science of Behavior Change Research Network, identifies neuropsychological targets that cut across particular diseases, and develops interventions related to those targets through an assessment of target engagement. The development of assays specific to those targets further allows the quantification of similarity across health conditions and intervention procedures, which is a necessary step to predict which interventions are replicable across varied contexts. Self-regulation is a promising neuropsychological target for facilitating behavior change across populations and contexts and the development of quantitative assays of this target may enable more efficient and cost-effective care.

Despite the promise of self-regulation as a mechanism of health behavior change, there are still significant impediments to this line of scientific inquiry. The primary obstacle is that self-regulation remains a poorly defined construct: intuitions about its nature are as diverse as the measurements used to study it. While some have argued for a single self-regulatory resource (Muraven & Baumeister, 2000), most researchers agree that it is not a unitary construct (Kotabe & Hofmann, 2015; Miyake & Friedman, 2012), though the field has yet to coalesce around a clear, specific definition of its components.

Moreover, even if self-regulation was adequately defined, there are disagreements about how these components relate to various behavioral, psychological and neural measurements. Meta-analytic work has shown that different self-regulation measurement approaches like self-report surveys and cognitive tasks have relatively low cross-method correlations (Duckworth & Kern, 2011), implying that they may capture different aspects of the construct. However, without further inquiry into the diversity of self-regulatory functions and their associated measurements, there is little to guide scientists towards a particular measurement strategy. In practice, a small subset of measurements at a single-level analysis (e.g. brain circuity measurements of a specific subdomain such as temporal discounting) are often used in any particular study as surrogates for self-regulation as a whole, without a clear characterization of their interrelationships and limitations. This results in potentially overgeneralized conclusions being drawn from measurements that may not fully capture the construct of self-regulation in its breadth and complexity.

Concepts related to subcomponents of self-regulation such as cognitive control and impulsivity have been plagued by similar definitional issues. To clarify the ambiguous nature of these constructs, previous efforts have employed data-driven approaches that seek to quantitatively divide these multidimensional constructs, either through meta-analysis or the analysis of individual differences. While meta-analysis can provide valuable insights, it is inherently limited by existing data. Sharma, Markon, & Clark (2013) used a meta-analytic principal-components factor analysis to decompose impulsivity measures into multiple factors (separately for self-report surveys and cognitive tasks). However, this analysis depended on a complete correlation matrix of impulsivity measurements and not all measurements have been paired together in the literature, necessitating imputation of missing information. Along with reducing confidence in the identified factor decompositions, their reduced data prevents straightforward analyses relating impulsivity measurements to real-world behaviors. These weaknesses are not intrinsic to the methods used in this particular study, but rather result from a general weakness of the literature: multiple measurements, particularly cognitive tasks, are rarely used in conjunction. This issue can only be addressed with an approach that deeply phenotypes each individual across a broad range of measurement devices spanning the domain of self-regulation and its real-world targets.

In comparison to meta-analysis, individual difference studies ensure that the entire measurement correlation matrix is obtained within the same population, making them a popular approach for scale development (DeVellis, 2016) and characterization of the latent structure of psychological functions. An impactful use of this method comes from cognitive psychology, where Miyake et al. (2000) identified a three-factor model of cognitive control using structural equation modeling (SEM). Moreover, they specified the relationships between these factors and the psychological tasks that best measure each construct. More recent work has similarly deconstructed impulsivity into five factors: stimulus interference, proactive interference, response interference, information sampling, and delay discounting (Stahl et al., 2014).

A limitation of previous studies in this domain is a lack of rigorous testing of predictive ability. A number of studies have assessed the relationship between self-regulatory ability and real-world outcomes, sometimes reporting impressively strong associations. However, the strength of an association between variables within a sample almost invariably overestimates the ability to predict outcomes in a new sample, due to overfitting of the model to the specific features of the dataset (Copas, 1983). A practical solution from machine learning that addresses this issue is cross-validation, in which the model is iteratively fitted to subsets of the data and then tested on the remaining held-out portion of the data (e.g., Hastie, Tibshirani, & Friedman, 2001).

Beyond proper assessment of predictive ability, previous approaches have also been limited by (1) their reliance on confirmatory structural equation modeling approaches, which potentially obfuscates useful latent structure, (2) the restricted scope of their measurements, (3) the lack of validation across different populations, and (4) their reliance on behavioral measures alone. Additionally, the resulting product of these approaches is a statistical model, which, while detailed, fails to serve as a common language for researchers moving forward. To compensate for this issue and aid interpretability, the model is often translated into a simplified verbal description of the components that constitute the larger domain (e.g. "shifting", "updating" and "inhibition" are parts of "cognitive control"; Miyake et al., 2000). While this verbal description then serves as a useful common language, much of the complexity of the original model is lost due to the limited expressivity of simply naming latent components.

A helpful middle ground between statistical models and verbal labels is explicitly constructing a formal ontology—a clear specification of the entities and relationships composing a domain (Bard & Rhee, 2004). Ontologies describe a domain using a formal description language with a restricted, precise vocabulary of relationships such as "is-a" or "kind-of". Psychological ontologies must also distinguish theoretical cognitive constructs from the tasks used to interrogate them, thus requiring a description of the measurement relationships between psychological tasks and the concepts they putatively measure (i.e., a "measured-by" relationship; Poldrack et al., 2011). The formal specification of these relationships affords ontologies more expressive power than a simple vocabulary of construct labels, and supports inference about data as well as assessment of the fit between data and conceptual models (Poldrack & Yarkoni, 2016).

The present paper outlines the design of a research program in the Science of Behavior Change network intended to extend the previous work on cognitive control and impulsivity to self-regulation more broadly, while addressing the limitations of previous approaches. We describe a procedure to develop a neuropsychological ontology of self-regulation, the structure of which delineates targets for behavioral change as well as their associated assays. Following this, we outline applications to two behavioral groups (daily smokers and overweight/obese adults who binge eat) and plans for intervention development. Ultimately, this project promises to enhance our understanding of self-regulatory function and as a consequence improve methods for effective behavioral change.

2. Overview

There are three main goals of this project. The first is the establishment of an ontology of self-regulation, informed by both behavioral and fMRI individual difference studies. The second validates the generalizability of this ontology in two exemplar behavioral groups: daily smokers and overweight/obese adults who binge eat, using cross-sectional and observational methods. The third validates the ontology in a second group of individuals with greater temporal resolution, using ecological momentary assessment (EMA) and mobile sensing, and seeks to use the ontology to inform ecological momentary interventions (EMI) to produce desirable changes in self-regulatory behavior. Each subsequent study builds on the results of earlier studies. In this report, we primarily focus on the first goal of ontology construction but also briefly describe planned next steps in this line of research.

3. Establishment of an ontology of self-regulation

This project's first objective is the development of an ontology of self-regulation. By this, we mean the specification of latent psychological constructs underlying the processes commonly referred to under the umbrella of self-regulation, and their relationships between each other and with the measurement tools used to interrogate them. To build such an ontology, we take a neuropsychological, individual-differences approach, first in a behavioral sample collected using Amazon's Mechanical Turk (MTurk: an online marketplace that can be used for psychological research), then in lab using a combination of fMRI and behavioral testing. These two studies are hereafter referred to as the MTurk study and the fMRI study, respectively. Unless otherwise noted, the following subsections pertain to the MTurk study.

The goal of this work was to measure the diversity of cognitive functions that allow an individual to engage in effective self-regulatory behaviors. "Cognitive functions" should be interpreted broadly, encompassing all mental functions including affective and motivational processes. For the MTurk study, we constructed a behavioral battery composed of 63 separate measurements: 37 cognitive tasks, 23 self-report surveys, and 3 separate surveys comprising a variety of demographic and lifestyle measurements. We take a relatively agnostic approach as to which cognitive functions are most important for self-regulation and thus designed this battery to extend beyond measurements traditionally used to interrogate self-regulation; however, the battery oversampled parts of the space intended to capture aspects of goal-planning, self-regulatory failure, impulsivity, cognitive control and temporal discounting, given their common associations in the literature with the self-regulation construct, and practical limitations concerning the length of the battery.

Our analysis plan mirrors the exploratory nature of our behavioral battery. We use a data-driven exploratory approach intended to provide

Table 1
Cognitive tasks.

Task	Individual Difference Variables	References
Adaptive N-Back	DDM Parameters Average load	(Harvey et al., 2005; Jaeggi, Buschkuehl, Jonides, & Perrig, 2008)
Angling Risk Task	Two Conditions (Keep, Release): Adjusted Clicks Loss Percent Score	(Pleskac, 2008)
Attention Network Task	DDM Parameters Alerting Effect Orient Effect Conflict Effect	(Fan, McCandliss, Fossella, Flombaum, & Posner, 2005)
Bickel Titrator	Discount Rate for three payout magnitudes	(Koffarnus & Bickel, 2014)
Choice Reaction Time	DDM Parameters	
Cognitive Reflection Task	Correct Proportion Intuitive Proportion	(Primi, Morsanyi, Chiesi, Donati, & Hamilton, 2016; Toplak, West, & Stanovich, 2014)
Columbia Card Task Cold/Hot	Average # of cards chosen Gain Sensitivity Loss Sensitivity # Loss Cards Sensitivity Level of Information Use	(Figner, Mackinlay, Wilkening, & Weber, 2009)
Dietary Decision Task	Health Sensitivity Taste Sensitivity	(Hare, Camerer, & Rangel, 2009)
Digit Span	Forward Span Reverse Span	(Woods et al., 2011)
Directed Forgetting	DDM Parameters Proactive Interference	(Nee, Jonides, & Berman, 2007)
Discount Titrator	Percent Patient	(Figner et al., 2010)
Dot Pattern Expectancy	DDM Parameters AY-BY BX-BY D-prime Bias	(Otto, Skatova, Madlon-Kay, & Daw, 2013)
Go-NoGo	D-prime Bias	
Hierarchical Learning Task	Score Attention to Hierarchy	(Badre, Kayser, & D'Esposito, 2010)
Holt & Laury	Percent Patient	(Holt & Laury, 2002)
Information Sampling Task	Two conditions (Decreasing Win, Fixed Win): Probability Correct at choice Motivation	(Clark, Robbins, Ersche, & Sahakian, 2006)
Keep Track Task	Score	(A. Miyake et al., 2000; Yntema, 1963)
Kirby	Discount Rate for three payout magnitudes	(Kirby & MarakoviĆ, 1996)
Local-Global	DDM Parameters Switch Cost Conflict Effect	(A. Miyake et al., 2000; Yntema, 1963)
Motor Selective Stop Signal	DDM Parameters SSRT Reactive Control Selective Proactive Control	(Aron, Behrens, Smith, Frank, & Poldrack, 2007)
Probabilistic Selection Task	Positive Learning Bias Value Sensitivity	(Frank, Seeberger, & O'Reilly, 2004)
Psychological Refractory Period	Slope of PRP function	(Pashler, 1994)
Raven's Matrices	Score	(Raven & Raven, 2003)
Recent Probes	DDM Parameters Proactive Interference	(Nee et al., 2007)
Shape Matching Task	DDM Parameters Stimulus Interference	(Stahl et al., 2014)

(continued on next page)

Table 1 (*continued*)

Task	Individual Difference Variables	References
Shift Task	Accuracy Learning Rate	(Wilson & Niv, 2012)
Simon Task	DDM Parameters Simon Effect	(Lu & Proctor, 1995)
Simple Reaction Time	Average Reaction Time	
Spatial Span	Forward Span Reverse Span	(Woods et al., 2011)
Stimulus Selective Stop Signal	DDM Parameters SSRT	(Bissett & Logan, 2014)
Stop Signal	DDM Parameters SSRT (low stop signal probability condition) SSRT (high stop signal probability condition) Proactive SSRT speeding Proactive Slowing	(Bissett & Logan, 2011)
Stroop	DDM Parameters Stroop Effect	(A. Miyake et al., 2000; Yntema, 1963)
Three-By-Two	DDM Parameters Stimulus Switch Cost (100/900 CSI*) Task Switch Cost (100/900 CSI*) * Cue-Stimulus Interval	(Schneider & Logan, 2011)
Tower of London	Average Move Time # Extra Moves # Optimal Solutions Planning Time	(Shallice, 1982)
Two-Step Decision	Model-Based Index Model-Free Index Perseverance	(Daw, Gershman, Seymour, Dayan, & Dolan, 2011)
Writing Task	Positive Probability Negative Probability	

useful, reduced representations of performance on the battery. We hold that no single representation of the data can be objectively validated (von Luxburg, Williamson, & Guyon, 2012) and thus must be evaluated in reference to particular goals (e.g., predictive ability, interpretability, parsimony between behavioral and neural data). It is an empirical question whether these different objectives produce qualitatively similar ontologies. A thorough overview of our analytical approach follows a more detailed description of our measurements and procedures. Study methods were pre-registered on the Open Science Framework (https://osf.io/amxpv/).

3.1. Cognitive tasks

To best capture the breadth of cognitive processes that may be relevant to self-regulation, we selected a broad range of cognitive tasks to be included in the behavioral battery. The full list of tasks is in Table 1; we selected tasks intended to measure the constructs of task-switching, response inhibition, proactive control, selective attention, working memory, risk-taking, impulsivity, planning, model-based versus model-free decision making, and information sampling, among others. All tasks were coded using the jsPsych javascript library (de Leeuw, 2015).

These putative cognitive descriptors were used to select tasks for our dataset, and previous literature was surveyed to inform the specific set of first-level analyses we planned to perform (e.g. the Stroop effect will be calculated for the Stroop task). Beyond these analyses, the ontology development process is data-driven, or *unsupervised*; that is, our dataset oversampled cognitive functions that we believed a priori to relate to self-regulation, but the ontology will ultimately be created from data-driven analyses that are agnostic to any a priori conceptual or empirical relationships between these functions.

Table 2
Self-report surveys.

Self-Report Surveys	Individual Difference Variables	References
BIS-11	Attentional Motor Non-Planning	(Patton, Stanford, & Barratt, 1995)
BIS-BAS	BAS Drive BAS Fun-Seeking BAS Reward-Responsiveness BIS	(Carver & White, 1994)
Brief Self-Control Scale	Self-Control	(Roth, Isquith, & Gioia, 2005)
Dickman's Impulsivity Inventory	Dysfunctional Functional	(Dickman, 1990)
DOSPERT (EB/RP/RT)	Ethical Financial Health/Safety Recreational Social	(Blais & Weber, 2006)
Three-Factor Eating Questionnaire (R18)	Cognitive Restraint Emotional Eating Uncontrolled Eating	(de Lauzon et al., 2004)
Emotion Regulation Questionnaire	Reappraisal Suppression	(Gross & John, 2003)
Five Facet Mindfulness Questionnaire	Acts with Awareness Describe Non-Judgment Non-Reactive Observe	(Baer, Smith, Hopkins, Krietemeyer, & Toney, 2006)
Future Time Perspective	Future-Time Perspective	(Carstensen & Lang, 1996)
Grit Scale	Grit	(Duckworth and Quinn, 2009)
Impulsive-Venturesome Survey	Impulsiveness Venturesomeness	(Eysenck, Pearson, Easting, & Allsopp, 1985)
Stanford Leisure-Time Activity Categorical Item (L-Cat)	Activity Level	(Kiernan et al., 2013)
Mindful Attention Awareness Scale	Mindfulness	(Brown & Ryan, 2003)
Multidimensional Personality Questionnaire (Control subscale)	Control	(Patrick, Curtin, & Tellegen, 2002)
Selection Optimization Compensation	Elective Selection Loss-based Selection Compensation Optimization	(Baltes, Baltes, Freund, & Lang, 1999)
Short Self-Regulation Survey	Control	(Carey, Neal, & Collins, 2004)
Sensation Seeking Survey	Boredom Susceptibility Disinhibition Experience Seeking Thrill/Adventure Seeking	(Zuckerman, 2007)
Ten Item Personality Questionnaire	Agreeableness Conscientiousness Emotional Stability Extraversion Openness	(Gosling, Rentfrow, & Swann, 2003)
Theories of Willpower	Endorse Limited Resource	(Job, Dweck, & Walton, 2010)
Time Perspective Survey	Future Past Negative Past Positive Present Fatalistic Present Hedonistic	(Zimbardo & Boyd, 2015)
UPPS + P	Lack of Perseverance Lack of Premeditation Negative Urgency Positive Urgency Sensation Seeking	(Lynam, Smith, Whiteside, & Cyders, 2006)

3.2. Self-report measures

Similar to our selection of the cognitive tasks, we selected a number of measures that were intended to oversample particular cognitive dimensions central to self-regulation while being broad enough to cover rarely-considered psychological functions potentially relevant to self-regulation. As a class, self-report surveys differ from tasks in the cognitive dimensions they capture, as well as the nature of their measurements (Duckworth & Kern, 2011). We took advantage of this fact to extend the psychological space encompassed by our battery. When possible, we attempted to include some measurements that share similar psychological descriptors between tasks and self-report surveys (particularly for central cognitive constructs like impulsivity), but were limited by the total length of the battery. The full list of self-report measures we used is in Table 2.

3.3. Demographics and target variables

Participants were asked to report on demographics (sex, age, race, ethnicity, height, weight), relationship characteristics (relationship status, divorce count, longest relationship, number of relationships, and number of children), financial status (household income, savings, debt), caffeine usage, gambling, traffic violations, arrests, and drug use (marijuana, nicotine, and others). We also asked about mental health using the K6 survey (Kessler et al., 2002) and about the participants' history of medical of neurological disorders. It should be noted that the self-reported nature of this data, as well as the specific content collected, limits the claims we can make about our participants. For example, the data are not sufficient to make strong claims about clinical phenotypes like obesity, given that we only have access to participants' self-reported height and weight (Nevill, Stewart, Olds, & Holder, 2006). Rather than contributing to the ontology, these variables were collected to capture demographic, mental health, or behavioral information that may relate to self-regulatory ability, and which may also serve to assess the concurrent validity of the resulting models.

3.4. Participants and procedure

Data collection for the self-reported surveys and tasks was divided into three phases: collection of a discovery cohort ($n = 200$), collection of a validation cohort ($n = 300$), and retesting of a subset of participants ($n = 150$). At this point, all phases of data collection have been completed. The discovery and validation cohorts were collected concurrently, with each participant randomly preassigned to one of the two cohorts. The retest subset was selected randomly from the discovery and validation cohort and completed the battery a second time. The battery required roughly 10 h to complete.

Due to the large number of subjects, and significant time involvement, we used Amazon's Mechanical Turk (MTurk) to collect behavioral data. In brief, MTurk is an online marketplace where workers can sign up to complete "HITs" (human intelligence tasks). In recent years, MTurk has proven a valuable source of diverse, easily accessible participants for psychological research, and has been shown to result in data of comparable quality to in-lab samples (Crump, McDonnell, & Gureckis, 2013). In the context of self-regulation, it also provides the ability to recruit a broader range of abilities compared to a sample from a university community. In contrast to most psychological studies on MTurk, which consist of a single relatively short testing session, the battery required multiple sessions due to its length. To address this issue, we developed the Experiment Factory (Sochat et al., 2016), a new infrastructure to deploy behavioral measurements on MTurk. The Experiment Factory presented tasks to participants in a random order, and allowed participants to complete the battery at their own pace, finishing as many or as few tasks as they wanted in each sitting. Participants were required to finish the entire battery within one week of accepting the HIT, but no other restriction was placed on their time. Only adults between 18 and 50 and living in the US were invited to participate, though four participants reported that their age was between 50 and 60. For completion of the battery, participants were paid \$60 plus bonuses averaging \$10 for their time (minimum: \$65, maximum: \$75).

As the behavioral battery was long (both in comparison to other psychology studies and MTurk HITs), reducing attrition was a significant consideration. In order to minimize attrition, a number of steps were taken, including providing comprehensive instructions, follow-up emails, and actively fielding questions on various online message boards for MTurk workers. Also, as an incentive to complete, we created a payment schedule that paid a lower rate if the participant failed to complete all 63 measures in the battery. Together, these steps kept attrition manageable: 84% of all participants who enrolled ultimately completed the entire battery. We are currently following-up with workers who failed to complete the entire battery. Our preliminary data analysis on the current dataset reports that their basic demographic information (e.g., age, gender, ethnicity) is not different from the larger population of individuals who completed the battery. We removed any participants who failed to complete the entire battery (102 out of 662), as well as any who failed to pass quality checks (see "Quality Checks for Cognitive Tasks") and continued recruiting until we achieved our sample size goal for each cohort. Due to over-recruiting to ensure we achieved minimum sample sizes, our final samples were 200 (discovery) and 322 (validation; all extra subjects beyond the planned sample were assigned to the validation set). Many of our analyses required larger sample sizes than either of these subsets; for these we combined both samples into a final cohort of 522 participants. Finally, completed participants were iteratively solicited to take the entire battery a second time, until 150 completed the battery while passing quality checks (see "Quality Checks for Cognitive Tasks" below). Completed participants were randomly ordered before solicitation, and all participants completed the retest within 4 months of the initial test (minimum 60 days, maximum 228 days gap between completions).

3.5. Quality Checks for Cognitive Tasks

Participants on MTurk are wholly unsupervised, necessitating procedures to ensure data quality. Quality checks were broadly applied to all cognitive tasks to ensure that (1) response times were not unreasonably fast on average, (2) omitted responses were reasonably low, (3) accuracy on cognitive tasks was reasonably high and (4) responses were sufficiently distributed (i.e. the participant didn't only press a single key). The specific criteria we used differed for some tasks, but in general we required that median response times were longer than 200 ms, no more than 25% of responses were omitted, accuracy was higher than 60% and no single response was given more than 95% of the time. These thresholds were set based on evaluation using the discovery cohort only, prior to unblinding the validation cohort. Overall, these steps were taken to ensure that participants in our dataset completed the tasks in earnest. Similar checks could not be performed on the self-report surveys or demographic measurements as we did not collect response time measures and potentially suspect response patterns (e.g. selecting only one response for every item) may be input honestly.

These criteria were used to evaluate each participant/task pair; failure on any check led to removal of that particular task's data for that participant. In addition, we removed a participant's entire dataset if they failed on four or more individual tasks (38 out of 560 participants were so removed).

These quality checks were intended as thresholds to screen out participants who were intentionally gaming the HIT. We also used task-specific manipulation checks which evaluated particular performance criteria specific to different tasks, necessary for the interpretability of our derived dependent measures. Failing these manipulation checks led to the removal of that participant's data on the failed task, but did not count towards the four failed tasks that would lead to the entire participant being removed from our study. The tasks that used these additional manipulation checks were the stop signal tasks, probabilistic selection task, and two-step decision task.

3.6. Overview of analytic approach

Each cognitive task and self-report measure was analyzed separately to derive a number of individual difference measures. Most of the measures of interest were determined using the discovery cohort, before the validation cohort was unblinded (see Tables 1 and 2). In total, we decomposed our cognitive measures into 125 distinct dependent variables, and our survey measures into 67 dependent variables. After determining the individual difference measures of interest for all cognitive tasks and self-report surveys, we created a participant-by-measure matrix which was the primary structure of interest for subsequent multivariate analyses. Below we outline the first-level analyses for cognitive tasks and self-report surveys, steps taken to evaluate measurement reliability, and our multivariate analytic approach.

3.6.1. Derivation of individual difference variables for cognitive tasks

As our cognitive tasks were heterogeneous, they each required separate analytic approaches to derive individual difference measures. The full list of cognitive tasks are listed in Table 1. Analysis code for each task can be found at https://github.com/IanEisenberg/expfactory-analysis/blob/master/expanalysis/experiments/jspsych_processing.py.

While a majority of our measures were thus specific to particular tasks, many of our cognitive tasks involved speeded two-alternative forced choice (2AFC) responses. As such, two primary metrics of interest for these tasks are accuracy and reaction time (RT). Accuracy and RT suffer from interpretability issues, as each depends in part on the other and they often trade off against each other (Wickelgren, 1977). This issue has been successfully dealt with in 2AFC tasks by employing the drift-diffusion model (DDM; Ratcliff, Smith, Brown, & McKoon, 2016). The DDM (in its simplest version) jointly models accuracy and RT as a function of three cognitively relevant latent parameters: (1) drift rate, (2) decision threshold and (3) non-decision time. Decomposing accuracy and RT in this way improves interpretability, as it cleanly separates individual differences in quality and speed of evidence accumulated towards a response (drift rate) from individual preferences along the speed-accuracy tradeoff curve (threshold).

For similar reasons, we use the DDM to create task contrasts of interest. For instance, the Stroop task's primary measure of interest is the degree of slowing on incongruent compared to congruent trials. Rather than define this measure explicitly using RT, we use the DDM where parameters can change as a function of condition. Thus, many dependent variables traditionally discussed in terms of accuracy or RT will be analyzed in terms of DDM parameter differences in this paper.

There are multiple existing methods for calculating DDM parameters, and their relative applicability are often debated. To reflect different perspectives in the DDM community, we employ two different methods: EZ-DDM and hierarchical-DDM (HDDM). The simpler method is EZ-DDM (Wagenmakers, van der Maas, & Grasman, 2007), a closed-form solution that directly transforms accuracy and RT measurements into DDM parameters. Its advantages extend beyond its simplicity, however, as some work suggests that it performs better in simulated comparisons (van Ravenzwaaij & Oberauer, 2009). In comparison, HDDM is a fitting procedure that models the DDM parameters hierarchically, such that individual parameters are assumed to be drawn from a group distribution (Wiecki, Sofer, & Frank, 2013). Work by Ratcliff and Childers (2015) suggests that HDDM is better able to capture true parameters when dealing with small datasets, or datasets corrupted by trials influenced by processes other than evidence accumulation (e.g., attentional lapses). Finally, the HDDM affords us a more flexible modeling system; for example, we can specify drift rate as a function of condition while keeping the other parameters constant, which may be justified for certain types of comparisons. Ultimately, we will select the appropriate modeling framework based on test-retest reliability (see “Reliability of Measures”) and the sensibility of the results given prior knowledge of these tasks.

3.6.2. Derivation of individual difference variables for self-report surveys

Self-report measures were analyzed in two ways. First, the measurements were scored in terms of subscales specific to each individual survey defined by previous literature. These subscale measures reflect the results of prior dimensional analyses in the existing research literature. In many cases, the literature on the scales was extensive, conducted over many years, and based on a wide variety of samples. As a result, each participant had a small number of scale scores for each questionnaire which reduced the number of measures substantially, compared to the entire list of items. This approach was complemented by our own analysis of descriptive statistics (mean, variance, skewness, kurtosis) for each item, and factor analysis and item response analysis for each scale.

The second approach analyzed data across all items without regard to the specific surveys from which they were selected. This analysis was motivated by the fact that analysis based on prior dimensional analysis uses potentially arbitrary boundaries between different surveys that measure similar constructs. This approach consisted of modeling all items within a single multi-dimensional model, without regard to subscales or surveys. Specifically, we used multi-dimensional item-response-theory (via the MIRT package in R; Chalmers et al., 2012) to explore the dimensionality of the survey data. Because of the size of the covariance matrix among items (594 items), this analysis is computationally intensive, and was accomplished using the Wrangler data analytics supercomputer at the Texas Advanced Computing Center. An alternative approach randomly divided the approximately 594 items into partitions, and analyses were conducted in each partition. The goal of the analysis of the scales and items was to identify consistent dimensions of the self-report scales. We are currently using the first method in our multivariate analyses (see “Overview of Multivariate Analytic Approach”), but will explore both approaches.

3.7. Data cleaning and imputation

To ensure we did not have redundant variables in the participant-by-measure data matrix, if any two individual difference variables derived from the same task or survey had a correlation $r > 0.85$, one of the variables was arbitrarily removed. In addition, because many of our analyses assume normally distributed variables, we transformed skewed variables and removed any variable that remained excessively skewed after transformation. Finally, our data matrix had missing values due to our quality check procedure. Many of our analyses require no missing data. When necessary, we imputed the data matrix using R's missForest package (Stekhoven & Buhlmann, 2012).

For the MIRT analyses, item responses were first cleaned in order to remove response options that received fewer than 40 choices; this was based on exploratory analyses which showed that model fitting often failed (resulting in likelihood values of zero) with values smaller than 40. Infrequent responses at the extreme end of the response spectrum were collapsed to the less extreme value (e.g., an infrequent “1” response on a 1–7 scale would be recoded as a “2”) until each response category had a frequency of at least 40 responses across all subjects. Items with non-extreme response categories having less than 40 responses were excluded from the analysis.

3.8. Reliability of individual difference variables

Measurement reliability is critically important in individual difference studies. While the psychometric properties of many of our self-report surveys have been documented in the literature, test-retest reliability is less commonly assessed for cognitive tasks. In addition, it is unclear how psychometric properties in the lab translate to online testing, and MTurk in particular. To address these issues, 150 subjects completed our battery a second time within 4 months of the initial test (minimum 60, maximum 228 days gap between completion).

We assessed test-retest reliability using both Spearman correlations as well as two-way fixed intraclass correlations. We used the former because it is less susceptible to non-normality in the data and the latter to account for the effect of individual differences. Because two-way fixed intraclass correlations do not take into account systematic error, however, we also calculated the effect size of time (first versus second measurement) for each of the dependent measures as well as the standard error of measurement. When appropriate, in subsequent analyses we use attenuation correction to account for the deleterious effect of measurement error (Muchinsky, 1996).

3.9. Overview of Multivariate Analytic Approach

The first-level analyses of the cognitive tasks and self-report surveys in the MTurk study result in a participant-by-measure matrix. Our fundamental goal is the identification of latent cognitive constructs underlying the observed covariability between measurements in this matrix. Due to inherent limitations of identifying the "correct" reduced representation of our data (von Luxburg et al., 2012), we are using multiple techniques whose convergence will inform the specifics of our ontology.

While measures derived from cognitive tasks and self-report surveys can, in principle, be analyzed within the same framework, in line with Sharma et al. (2013) and Duckworth and Kern (2011), we expect a general lack of shared variance between cognitive tasks and self-report surveys. A lack of relationship would likely result both from gross measurement differences between tasks and surveys (e.g., behavioral requirements, psychometric properties) as well as differential relationships between the two measurement groups and underlying psychological constructs. With that said, the approaches described below could be performed using the entire dataset, or separately on either the cognitive tasks or the self-report survey measures.

To evaluate the approximate dimensionality of our space we are using exploratory factor analysis (EFA). This technique explicitly defines latent components that can be interpreted based on their relationship to observed variables. EFA requires the specification of a desired dimensionality; there are many proposed methods for choosing the optimal dimensionality, but we selected the Bayesian Information Criterion employed by the Psych R package (Revelle, 2017).

Because the specification of a particular dimensionality is ultimately arbitrary, we will explorea range of dimensionalities surrounding the optimal parameter and quantified the relationships between components derived from different dimensionalities. This approach shares some commonalities with hierarchical clustering techniques, in that the relationships between measures can be understood with varied levels of precision, but our technique does not enforce explicit hierarchies.

To complement this approach we are also drawing on graph theoretic concepts (Epskamp, Borsboom, & Fried, 2016). Graph theory provides a rich language to describe the relative importance and relationships between network nodes. In our network, nodes correspond to unique individual difference variables, and edges correspond to a measure of relationship, either signed (e.g. Spearman's correlation) or unsigned (e.g. distance correlation). Such a network can be clustered, again providing a description of latent constructs composed of multiple measurements.

3.10. Associations with target variables

One important aspect of ontology development is the assessment of concurrent validity with regard to associations with characteristics of interest measured at the same point in time. Given the breadth of demographics, health behaviors and other target measures we collected, this project can evaluate and qualify the concurrent validity of many psychological measures. Additionally, predictive analyses qualify the cognitive ontology by contextualizing the measures we use in relationship to real-world behaviors. Simply put, the ontological structure is only useful for experimental medicine approaches if the latent constructs meaningfully relate to target behaviors of interest. In this work, we prioritize risky behaviors involving overeating, alcohol, smoking and drug-use. We do this both due to the particular societal cost of these behaviors, and as a resource-constrained proof-of-concept. If the ontology proves useful in these domains, our hope is for future research to contextualize other risky behaviors in the same framework. Caveats are that all of our target measures are self-reported at one point in time on MTurk. We did not directly measure any "real-world outcomes" or changes over time, and thus the measures may be biased in a number of ways (see "Validation in a Mobile Assessment Paradigm" for additional details).

Our predictive approach can be broadly separated into "unstructured" and "structure-based" approaches. Unstructured approaches make no use of the ontology we derive, or any other reduced representation of the data (e.g. a thresholded graph). Instead, we make use of cross-validated regularized regression, an approach that finds the optimal linear combination of psychological variables to predict target variables, with constraints placed on model parameters to reduce model variance (e.g., Hastie et al., 2001). The use of regularized linear methods aids interpretability, and is likely to improve generalizability, but we will also assess performance of non-linear methods such as random forests.

"Structure-based" methods make use of the learned structure to either (1) transform and/or reduce the features used for prediction or (2) enforce a regularization informed by structure rather than generic methods like L1/L2 regularization. An example of the former would be using exploratory factor analysis to derive latent components which are used for estimation, while an example of the latter is graph-regularized regression (Li & Li, 2008).

3.11. Informing the Ontology with fMRI

We will supplement the outlined MTurk study with an fMRI individual differences study. As of the time of publication, the MTurk phase has been used to select tasks and surveys for use in the fMRI study, and fMRI data collection is ongoing. Moving forward, neuroimaging will complement behavioral data from the MTurk study, and contribute to a multi-level neuropsychological ontology of self-regulation.

100 healthy participants, aged 18–40, will participate in two 1.5-h fMRI sessions. After scanning, they will complete the same online battery that our MTurk participants completed. Across the two scanning sessions, participants will complete 16 min of rest, 9 cognitive tasks and 40 survey questions during fMRI acquisition. Participants will be compensated $80 for the fMRI sessions, and $100 for the online battery. In addition they can earn between $10 and $20 bonus payment for performance on some tasks completed either during the fMRI or online. Participants will be recruited from the area surrounding Stanford University. To increase the demographic diversity of our sample, we will use multiple recruiting methods, including a Stanford University recruitment portal (SONA), online ads (e.g. Craigslist) and flyering in neighboring communities.

3.11.1. Task selection for the fMRI study

Given our time constraints, we are unable to scan the entire set of tasks from our behavioral battery. As such, we created a selection procedure using the discovery cohort from the MTurk phase to inform our task selection, which resulted in a subset of cognitive tasks to be used with fMRI. To accomplish this, we initially evaluated task subsets based on their ability to reconstruct the entire variability of our task variables, survey variables, and demographic variables using cross-validated multiple-linear regression. Reconstruction of survey variables and demographic variables on the basis of task variables was generally poor ($R^2 < 0.04$), so we reduced our objective function to picking task subsets that best reconstructed the entire task matrix with the constraint that the task subset must fit within our time constraints.

We first reduced the task data using principal components analysis, with cross-validation to identify the optimal number of components. We then used L1-regularized linear regression to reconstruct these principal components as linear combinations of the individual difference variables in a proposed task subset. Task reconstruction success was evaluated via a k-fold cross-validation procedure, with the cost function defined by the correlation between reconstructed principal components and the actual principal components using the held-out data, with the aim to maximize this correlation. Given this objective function and the constraint that task subsets must fit within our time constraints, we used a genetic algorithm to arrive at an optimal task subset. In brief, the genetic algorithm creates a population of "individuals" (task subsets) who are selected for cross-pollination based on their fitness (defined by the objective function). Mutations, along with the introduction of new individuals contribute to the task subset variability. Together, these features allow a genetic algorithm (Whitley, 1994) to efficiently search large spaces to find solutions that optimize multiple objectives.

This procedure arrived at the following task subset: attention network task, Bickel titrator, Kirby discounting task, Columbia card task (cold/hot), dot pattern expectancy, motor-selective stop signal, Stroop, three-by-two, and Tower of London. We then evaluated each task for its viability in fMRI and adapted these tasks as necessary to create a final task list. These adaptations led to an overly lengthy task list, so we consolidated tasks that were similar. This final evaluation led to the following changes: the "orienting" conditions were removed from the attention network task, the Bickel titrator and Kirby tasks were replaced by a separate delay discounting task (J. Kable, personal communications, February 8th, 2017), the Columbia card tasks (cold/hot) were replaced by an fMRI adaptation of the Columbia card task (fmri-CCT: van Duijvenvoorde et al., 2015), the three-by-two was replaced by a two-by-two task, the Tower of London was replaced by a Ward and Allport tower task (WATT3: Kaller, Rahm, Spreer, Weiller, & Unterrainer, 2011), and the motor-selective stop signal was separated into motor-selective stop signal and a stop signal task (without increasing the combined length). These changes resulted in the final task list: attention network task, fmri-CCT, delay discounting task, dot pattern expectancy, motor-selective stop signal, stop signal, Stroop, two-by-two, and the WATT3.

3.11.2. Survey item selection for the fMRI study

We used two primary approaches to select items from our self-report surveys to be presented during fMRI scanning. First, we selected several complete scale measures that were short, had good results from initial dimensional analyses in the MTurk phase, and had considerable background literature: the Brief Self-Control Scale (13 items), Grit Scale (8 items), and Carstensen Future Time Perspective (10 items). Second, we selected items that loaded on similar factors in the MIRT dimensional analysis (see "Self-Report Measures") of the entire list of items, which resulted in the addition of six items from the UPPS + P impulsivity measures and three items from the Impulsiveness-Venturesomeness scale.

4. Validation in behavioral groups in lab-based paradigm

The above investigation of self-regulation—at behavioral, psychological and neural levels of analyses—is only the first phase of a multiphase project. One subsequent phase examines self-regulatory function in two exemplar behavioral groups for which self-regulatory failure putatively plays a critical role in health outcomes: daily smokers and overweight/obese adults who binge eat. We will briefly discuss the proposed experimental approach and aims for this future work.

We intend to recruit a sample of 50 daily smokers and 50 overweight/obese individuals who binge eat for a combination of fMRI and subsequent behavioral testing. For fMRI, we will complete a rest scan and will select a subset of the fMRI measures completed by the previously tested population (9 cognitive tasks, and 40 survey questions, see "Informing the Ontology with fMRI") using a similar genetic algorithm approach to arrive at an optimal task subset. We may interrogate self-regulation and emotion regulation in the midst of smoking cues in smokers and food cues in overweight/obese adults who binge eat.

There are three main aims for this phase of the study. First, extending our ontology to these two groups will test the generality of our ontology of self-regulation. A fundamental assumption of this project is that a cognitive ontology is a relatively generalizable structure that covers a broad space of cognitive functions. Under this assumption the ontology itself should not change in different contexts or populations—just the relative expression of particular components. Tests of the measures with smokers and overweight/obese individuals who binge eat provide a validation set for testing this assumption.

Second, by interrogating several key aspects of self-regulation in these groups, we aim to uncover a more complete understanding of the relationship between self-regulation and both smoking behavior and binge eating. Much of the preceding work in this area has used individual tasks and measures, which produces siloed work that does not easily support synthesis between research projects (Vohs & Baumeister, 2016). Smoking and binge eating are complex behaviors that are determined by multiple environmental, physiological, emotional, and behavioral processes. By combining behavior and neuroimaging across a suite of tasks, we aim to develop a richer model of the role of self-regulatory processes associated with smoking and binge eating. This work will position us well to intervene on specific aspects of self-regulation that are particularly problematic in each group (see *Developing and Validating Momentary Self-Regulation Assessments* section below).

Third, as we collect data from all participants, we intend to include manipulations to modulate putative targets within the self-regulation domain in each clinical group, thereby assessing the extent to which we can shift self-regulatory function both in desired and undesired directions. The details of this procedure are currently being formulated, but we may (1) expose subjects to specific stimulus sets relevant to the sample that may promote engagement of appetitive drives (tobacco-related images for smokers or images of highly palatable foods for overweight/obese individuals who binge eat), and (2) expose them to an instructional manipulation designed to engage self-regulatory processes in the presence of these stimulus sets. These instructional manipulations may involve regulation or suppression of urges to the appetitive stimuli.

5. Developing and Validating Momentary Self-Regulation assessments

Most self-report measures of self-regulation address trait-like constructs like impulsivity, despite the construct's responsiveness to dynamic environmental, personal and social cues. As a result, these measures ask participants to report dispositional individual differences, rather than reflecting a momentary level of regulatory shifts, dynamics, and states that could change over time.

Based on the 594 self-regulation survey items from the first MTurk phase, we aim to develop and validate modified candidate items of self-regulation. We will examine their psychometric properties to finalize a momentary self-regulation scale that captures self-regulation processes in a naturalistic setting for our later EMA phase. In the following sections, we describe the analyses undertaken to identify candidate survey items for EMA and an on-going pilot study that is designed to validate the modified candidate items of self-regulation.

5.1. Identifying measures for EMA

In the MTurk phase, data on 594 self-reported survey items were collected. Using this item-level data from participants in the discovery sample, item response theory was used to select a reduced set of items on which to evaluate the dimensionality of the survey items. The IRT analysis was performed within each literature-defined subscale, and three items from each subscale were selected based on their item information and item characteristic curves. The reduced item set consisted of 116 items.

Using the reduced set of items, several EFAs extracting different numbers of correlated factors were performed on the discovery, validation, and complete samples. Based on the summary of evidence from these analyses, three factors appeared to be present in the survey item space: perseverance and impulsivity, sensation seeking, and mindfulness/acting with awareness.

Combining information from the three-factor EFA solution, interpretation of the item text, and the intraclass correlation (ICC) from the test-retest sample, we further reduced the number of items to pilot test for momentary variability with the goal of identifying 20 items to pilot.

These selected items provided the most convergent and discriminant evidence (e.g., item discriminant coefficient > 2, factor loading scores > 0.5, excluding ambiguous items that load on two factors). This process combined both quantitative and qualitative information about the items and the constructs they measured. Specifically, within each factor, we selected items that loaded highly on the factor in the three-factor EFA, did not refer to a specific task so therefore would not be likely to be valuable on a momentary basis (e.g. "I would like to try surfboarding"), and showed some variability at the individual level. We selected more items from the first factor (perseverance and impulsivity) because in the EFA, many more items loaded on this factor than the other two. To identify items with within-individual variability, we examined ICCs from the test-retest sample. Items with very high ICC indicate little within-individual variability and thus were not selected. Once the initial reduced set of items was identified, a confirmatory factor analysis was performed in the complete sample using the reduced item set of 20. This confirmatory analysis did not show good fit for a three-factor solution and instead had better fit from a four-factor solution in which the third factor was comprised of two factors: emotion regulation and mindfulness/act with awareness. Given this, we selected additional items within the third factor that, based on item text, appeared to measure these two distinct constructs, and to preserve the total number of items, we eliminated some items measuring the first factor. This set of 20 momentary items will be pilot tested in a sample of MTurk workers.

5.1.1. Pilot test of EMA candidate items

A pilot study is now underway to evaluate the extent to which measures of self-regulation can be adapted to provide a momentary or state measure. This study is important for at least two major reasons. First, if the dimensions are stable traits, the measure will have minimal change over time. Second, it is ideal to have measures of important momentary self-regulation constructs that could be realistically changed in an intervention (Baraldi, Wurpts, MacKinnon, & Lockhart, 2014).

Throughout the pilot research activities, we will develop a brief, reliable, and psychometrically sound self-regulation momentary scale that assesses self-regulation shifts and dynamics and captures inter- and intra-individual variability in self-regulation. Using the 20 candidate items identified from our MTurk phase (as described above), we aim to implement and collect 14-day micro-surveys through text-message prompts among 50 adults to collect 42 timepoints per participant of repeated self-regulation survey data (3 times per day x 14 days = 42 timepoints). Recruitment and compensation activities will take place on MTurk. The team will conduct psychometric tests on momentary self-regulation items to finalize the best set of momentary self-regulation items for the EMA phase. Research outcomes from this project will provide a brief set of momentary self-regulation assessments that can be implemented and administered through mobile devices for EMA.

5.2. Validation in a Mobile Assessment Paradigm

We will employ the resulting momentary self-regulation measures in a subsequent phase of research which aims to examine self-regulatory function in 50 adult smokers and 50 overweight/obese individuals who binge eat. Unlike the previously-described phase of research which will employ lab-based fMRI methods of data collection, in this phase we will measure self-regulation via ecological momentary assessment (EMA) and mobile sensing as participants move through their daily lives. EMA is a methodology that prompts individuals to respond to specific queries on mobile devices (Shiffman, Stone, & Hufford, 2008). The frequent, longitudinal assessment afforded by technology-based monitoring of behavior and self-regulatory function in naturalistic contexts may help clarify their interrelationships as well as reveal an increased understanding of the role of self-regulation in changing behavior.

In addition to data collected from participants via EMA, we will obtain data from sensors embedded on a smartphone and a smartwatch to enhance our understanding of the extent to which we can engage and manipulate self-regulation. To this end, we will partner with the Mobile Sensor Data-to-Knowledge (MD2K) "Big Data Center of Excellence" funded by the National Institutes of Health (NIH) and leverage their mobile sensing architecture to collect in real time raw sensor data to infer smoking, eating, stress, activity and sleep (md2k.org).

Including passive behavioral sensing will complement the self-reported self-regulation, context, and behavior data collected via EMA in multiple ways. First, obtaining more nuance and detail regarding behavior via passive sensing will enable better characterization of the contexts (e.g., time of day, location, alone/with others) in which self-regulation will be assessed. Second, the passively sensed data provide a way to validate self-reported information for those behaviors that are queried both by self-report and passive sensing. Although EMA is helpful in collecting data about an individual's perception of their mood, context and behavior, the data resulting from that method of data collection can be biased by an individual's lack of self-awareness, bias in self-reporting and/or non-response to queries due to participant burden or fatigue. Importantly, if an individual does not report an unhealthy behavior (smoking or binge eating), this will be captured passively. Additionally, because there may be a relationship between self-regulation and engagement with the mobile platform, it is important to understand contexts and behaviors associated with not providing EMA responses to queries about context, behaviors, and putative self-regulation targets. The addition of passively sensed behavioral and contextual features will enable an examination of this issue.

5.3. Mobile behavior change intervention approach

Unlike the paradigm to be used in the lab, we will evaluate (via mobile assessment) antecedent conditions prior to risk behavior (smoking or binge eating among our samples). This will allow us to better understand real-world conditions that may engage and modulate putative targets of self-regulatory function in undesired directions. This function enables "automated hovering," or real-time monitoring, of individuals' behavior (Asch, Muller, & Volpp, 2012).

Also, unlike the instructional control interventions to be used in the lab setting, we will offer the resources of an evidence-based mobile behavior change intervention to individuals and assess the effect of this mobile intervention system on self-regulatory function. This measure, Laddr® (Square2 System, Inc.), offers science-based self-regulation monitoring and behavior change tools via an integrated platform to a wide array of populations. This system is based on approximately 17 years of NIH-supported research by our group on the science of behavior change and digital health technology (www.c4tbh.org) and embraces the commonalities in the principles of effective behavior change, including behavioral activation, problem-solving therapy, motivational enhancement, and cognitive behavioral therapy for behavior change. This system enables frequent, longitudinal assessment in naturalistic contexts, offers science-based self-regulation behavior change tools of relevance to an array of populations (including the targeted samples in this project), and enables the objective monitoring of health behavior (including smoking and exercise behavior).

We will ask each participant to engage with the mobile intervention to complete goal-setting/tracking activities and complete interactive self-regulation guides(on smoking reduction/cessation or diet/exercise) for two weeks. We will prompt them via EMA on their mobile device multiple times throughout each day to inquire about health risk behavior and ask them to complete momentary measures of self-regulation. They can also initiate use at any time. We will compensate participants to use the smartwatch and smartphone and to use Laddr daily to, at a minimum, update progress toward goals and complete at least two interactive self-regulation guides.

5.4. Data collection and analyses

The EMA analyses will use a variety of multilevel models that account for the correlated nature of the repeated assessments within person and can incorporate variables at the moment-, day-, or individual-level (Fitzmaurice, Laird, & Ware, 2012). For analyses of unhealthy behavior events, we will also model the pattern of data surrounding the event (e.g., participant mood, context, time of day, etc) via flexible spline models. We will model both main effects of real-world contexts as well as interaction effects with the intervention. The multilevel models also allow comparisons of within-individual moments during which challenges with real-world contextual factors were and were not followed directly by use of the mobile intervention as we assess the impact of intervention exposure on self-regulation measures. With 50 participants per sample, each with multiple assessments per day, the sample size for such analyses will be adequate to model these processes. While calculating exact detectable effect sizes will depend on several as-yet-determined factors (standard deviation [SD] of self-regulation measures, proportion of signals in challenging versus not-challenging contexts, response rate, intraclass correlation of EMA responses within individual), we estimate 80% power to detect a difference between contexts in mean self-regulation target value that is 0.2–0.3 times a SD. This assumes two signals per day for two weeks at an 80% signal response rate, with intraclass correlations ranging from 0.001 to 0.3, and for both 10% and 30% challenging context prevalence at signal time.

EMA and passive sensing data are necessary for modeling fine-grained, dynamic processes, as effects of real-world stimuli on putative self-regulatory targets are expected on a momentary basis. Some of the planned mechanistic modeling will be performed via structural equation models (SEM), which can incorporate both complex relationships between variables (such as mediation) and latent variables. As self-regulation will be measured via multiple self-reported items on the EMA, it can be considered a latent construct in the mechanistic models. Likewise, having both self-report and passively sensed information about the behavior will allow us to better model the true, unobserved unhealthy behavior as a latent variable in the relationship between context, behavior, and self-regulation, increasing the accuracy of the estimates of the relationships.

Flexible spline models that examine patterns of target measures around smoking and eating will be performed on all events, whether detected via self-report or passive sensing. This will allow us to both examine patterns of self-regulation around these events and determine whether this pattern differs when the event was reported versus not reported. Additionally, a validation of self-reported smoking instances will be performed, as smoking can be passively sensed via the wristband. A similar validation of instances of eating is also possible. Precision and recall percentages can be generated for each mode of data collection (passive sensing and self-report) using the other as ground truth. It is likely that both modes represent measurement of true behavior with error and such validation exploration will inform future data collection in each mode. Finally, we will analyze participant behavior during times when they did not respond to an EMA prompt and explore how these behavioral patterns are distinct from times when participants did respond We aim to better understand which behaviors and contexts are related to non-response to EMA prompts and to monetary self-regulation data. This will be accomplished in two steps. First, behavioral features will be entered into machine learning algorithms trained on the labeled observations when both self-report and sensor data is available, with the goal of predicting self-reported self-regulation measures. Next, the resultant optimal model will be used to predict the values of self-regulation measures during the spans of time when participants do not self-report. If a predicted measure during the times when participants do not respond to EMA is significantly different than those when the participant is responsive when prompted, this would further validate that measure as a potentially useful metric of self-regulation, as disengagement from Laddr and data collection are likely related to self-regulation.

6. Discussion

This program of research advances our understanding of self-regulatory processes by developing an ontology of relevant psychological functions and related measurements, informed by behavioral, psychological and neural data. Our large-scale individual-differences approach promises to extend our theoretical understanding of the structure of self-regulation and relate those functions to personal characteristics, mental health, and risk behaviors. To explore the relevance of these constructs as targets for intervention, we will follow up our ontology development with smokers and overweight/obese binge-eating individuals recruited from the community. Beyond the specific relevance of self-regulatory function in these two groups, this work will extend our understanding of the generalizability of the ontological structure of self-regulation. The end result of this multi-level, multi-population approach is a holistic account of self-regulatory function that extends beyond any particular scientific or environmental context.

A well-defined ontological description of self-regulation is significant for the science of behavior change as well. Interventions are currently developed with regard to particular behaviors: smoking, medical non-compliance, overeating. This work will provide an alternative: interventions created to address particular cognitive functions related to the problematic “lapses of self-regulation.” People could be described by an “ontological phenotype” that specifies their particular pattern of deficits in self-regulation (be it due to genetic background, experience, environmental exposures or cross-causal interactions), which could then be leveraged for personalized interventions. The final aim of this proposal intends to do just this in real time using ecological momentary interventions.

Funding

This work was supported by the National Institutes of Health (NIH) Science of Behavior Change Common Fund Program through an award administered by the National Institute for Drug Abuse (NIDA) (UH2DA041713; PIs: Marsch, LA & Poldrack, RA). Additional support was provided by NIDA P30DA029926.

Acknowledgements

Financial Conflict of Interest: In addition to her academic affiliation, Dr. Marsch is affiliated with Square2 Systems, Inc., the small business that developed the mobile Laddr intervention. Dr. Marsch has worked extensively with relevant institutions (including Dartmouth

College and NIH) to manage any potential conflict of interest. All research data collection, data management, and statistical analyses are conducted by individuals with no affiliation with Square2 Systems.

The findings and conclusions in this report are those of the authors and do not necessarily represent the views of the National Institutes of Health.

References

Aron, A. R., Behrens, T. E., Smith, S., Frank, M. J., & Poldrack, R. A. (2007). Triangulating a cognitive control network using diffusion-weighted magnetic resonance imaging (MRI) and functional MRI. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience, 27*(14), 3743–3752.
Asch, D. A., Muller, R. W., & Volpp, K. G. (2012). Automated hovering in health care—watching over the 5000 hours. *The New England Journal of Medicine, 367*(1), 1–3.
Badre, D., Kayser, A. S., & D'Esposito, M. (2010). Frontal cortex and the discovery of abstract action rules. *Neuron, 66*(2), 315–326.
Baer, R. A., Smith, G. T., Hopkins, J., Krietemeyer, J., & Toney, L. (2006). Using self-report assessment methods to explore facets of mindfulness. *Assessment, 13*(1), 27–45.
Baltes, P. B., Baltes, M. M., Freund, A. M., & Lang, F. R. (1999). *The measure of selection, optimization, and compensation (SOC) by self-report*Berlin: Max Planck Institute for Human Development.
Baraldi, A. N., Wurpts, I. C., MacKinnon, D. P., & Lockhart, G. (2014). Evaluating mechanisms of behavior change to inform and evaluate technology-based interventions. In L. Marsch, S. Lord, & J. Dallery (Eds.). *Behavioral health care and technology* (pp. 187–199). Oxford University Press.
Bard, J. B. L., & Rhee, S. Y. (2004). Ontologies in biology: Design, applications and future challenges. *Nature Reviews Genetics, 5*(3), 213–222.
Bickel, W. K., Odum, A. L., & Madden, G. J. (1999). Impulsivity and cigarette smoking: Delay discounting in current, never, and ex-smokers. *Psychopharmacology, 146*(4), 447–454.
Bissett, P. G., & Logan, G. D. (2011). Balancing cognitive demands: Control adjustments in the stop-signal paradigm. *Journal of Experimental Psychology. Learning, Memory, and Cognition, 37*(2), 392–404.
Bissett, P. G., & Logan, G. D. (2014). Selective stopping? Maybe not. *Journal of Experimental Psychology. General, 143*(1), 455.
Blais, A.-R., & Weber, E. U. (2006). *A domain-specific risk-taking (DOSPERT) scale for adult populations.* Retrieved from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1301089.
Brown, K. W., & Ryan, R. M. (2003). The benefits of being present: Mindfulness and its role in psychological well-being. *Journal of Personality and Social Psychology, 84*(4), 822.
Carey, K. B., Neal, D. J., & Collins, S. E. (2004). A psychometric analysis of the self-regulation questionnaire. *Addictive Behaviors, 29*(2), 253–260.
Carstensen, L. L., & Lang, F. R. (1996). *Future time perspective scale.* Unpublished ManuscriptStanford University.
Carver, C. S., & White, T. L. (1994). Behavioral inhibition, behavioral activation, and affective responses to impending reward and punishment: The BIS/BAS Scales. *Journal of Personality and Social Psychology, 67*(2), 319.
Chalmers, R. P. (2012). mirt: A multidimensional item response theory package for the R environment. *Journal of Statistical Software, 48*(6), 1–29.
Citation, P., Whiteside, S. P., & Lynam, D. R. (2001). UPPS impulsive behavior scale the five factor model and impulsivity: Using a structural model of personality to understand impulsivity. *Personality and Individual Differences, 30*(1000), 669–689.
Clark, L., Robbins, T. W., Ersche, K. D., & Sahakian, B. J. (2006). Reflection impulsivity in current and former substance users. *Biological Psychiatry, 60*(5), 515–522.
Copas, J. B. (1983). Regression, prediction and shrinkage. *Journal of the Royal Statistical Society. Series B, Statistical Methodology,* 311–354.
Crump, M. J. C., McDonnell, J. V., & Gureckis, T. M. (2013). Evaluating Amazon's mechanical Turk as a tool for experimental behavioral research. *PloS One, 8*(3), e57410.
Daw, N. D., Gershman, S. J., Seymour, B., Dayan, P., & Dolan, R. J. (2011). Model-based influences on humans' choices and striatal prediction errors. *Neuron, 69*(6), 1204–1215.
DeVellis, R. F. (2016). *Scale development: Theory and applications.* SAGE Publications.
Dickman, S. J. (1990). Functional and dysfunctional impulsivity: Personality and cognitive correlates. *Journal of Personality and Social Psychology, 58*(1), 95–102.
Duckworth, A. L., & Kern, M. L. (2011). A meta-analysis of the convergent validity of self-control measures. *Journal of Research in Personality, 45*(3), 259–268.
Duckworth, A. L., & Quinn, P. D. (2009). Development and validation of the short grit scale (grit – S). *Journal of Personality Assessment, 91*(2), 166–174.
van Duijvenvoorde, A. C. K., Huizenga, H. M., Somerville, L. H., Delgado, M. R., Powers, A., Weeda, W. D., ... Figner, B. (2015). Neural correlates of expected risks and returns in risky choice across development. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience, 35*(4), 1549–1560.
England, L. J., Aagaard, K., Bloch, M., Conway, K., Cosgrove, K., Grana, R., ... Wakschlag, L. (2017). Developmental toxicity of nicotine: A transdisciplinary synthesis and implications for emerging tobacco products. *Neuroscience and Biobehavioral Reviews, 72,* 176–189.
Epskamp, S., Borsboom, D., & Fried, E. I. (2016). *Estimating psychological networks and their stability: A tutorial paper.* arXiv Preprint arXiv:1604. 08462. Retrieved from http://adsabs.harvard.edu/abs/2016arXiv160408462E.
Epstein, L. H., Salvy, S. J., Carr, K. A., Dearing, K. K., & Bickel, W. K. (2010). Food reinforcement, delay discounting and obesity. *Physiology & Behavior, 100*(5), 438–445.
Eysenck, S. B. G., Pearson, P. R., Easting, G., & Allsopp, J. F. (1985). Age norms for impulsiveness, venturesomeness and empathy in adults. *Personality and Individual Differences, 6*(5), 613–619.
Fan, J., McCandliss, B. D., Fossella, J., Flombaum, J. I., & Posner, M. I. (2005). The activation of attentional networks. *NeuroImage, 26*(2), 471–479.
Figner, B., Knoch, D., Johnson, E. J., Krosch, A. R., Lisanby, S. H., Fehr, E., et al. (2010). Lateral prefrontal cortex and self-control in intertemporal choice. *Nature Neuroscience, 13*(5), 538–539.
Figner, B., Mackinlay, R. J., Wilkening, F., & Weber, E. U. (2009). Affective and deliberative processes in risky choice: Age differences in risk taking in the Columbia card task. Journal of experimental psychology. *Learning, Memory, and Cognition, 35*(3), 709–730.
Fitzmaurice, G. M., Laird, N. M., & Ware, J. H. (2012). *Applied longitudinal analysis.* John Wiley & Sons.
Frank, M. J., Seeberger, L. C., & O'Reilly, R. C. (2004). By carrot or by Stick: Cognitive reinforcement learning in parkinsonism. *Science, 306*(5703), 1940–1943.
Gosling, S. D., Rentfrow, P. J., & Swann, W. B., Jr. (2003). A very brief measure of the Big-Five personality domains. *Journal of Research in Personality, 37*(6), 504–528.
Gross, J. J., & John, O. P. (2003). Individual differences in two emotion regulation processes: Implications for affect, relationships, and well-being. *Journal of Personality and Social Psychology, 85*(2), 348–362.
Hare, T. A., Camerer, C. F., & Rangel, A. (2009). Self-control in decision-making involves modulation of the vmPFC valuation system. *Science, 324*(5927), 646–648.
Harvey, P.-O., Fossati, P., Pochon, J.-B., Levy, R., LeBastard, G., Lehéricy, S., ... Dubois, B. (2005). Cognitive control and brain resources in major depression: An fMRI study using the n-back task. *NeuroImage, 26*(3), 860–869.
Hastie, T., Tibshirani, R., & Friedman, J. (2001). *The elements of statistical learning.* NY Springer.
Holt, C. A., & Laury, S. K. (2002). Risk aversion and incentive effects. *The American Economic Review, 92*(5), 1644–1655.
Jaeggi, S. M., Buschkuehl, M., Jonides, J., & Perrig, W. J. (2008). Improving fluid intelligence with training on working memory. *Proceedings of the National Academy of Sciences, 105*(19), 6829–6833.
Job, V., Dweck, C. S., & Walton, G. M. (2010). Ego depletion—is it all in your head? *Psychological Science, 21*(11), 1686–1693.
Kaller, C. P., Rahm, B., Spreer, J., Weiller, C., & Unterrainer, J. M. (2011). Dissociable contributions of left and right dorsolateral prefrontal cortex in planning. *Cerebral Cortex, 21*(2), 307–317.
Kessler, R. C., Andrews, G., Colpe, L. J., Hiripi, E., Mroczek, D. K., Normand, S.-L. T., ... Zaslavsky, A. M. (2002). Short screening scales to monitor population prevalences and trends in non-specific psychological distress. *Psychological Medicine, 32*, 959–976.
Kiernan, M., Schoffman, D. E., Lee, K., Brown, S. D., Fair, J. M., Perri, M. G., et al. (2013). The Stanford leisure-time activity categorical item (L-Cat): A single categorical item sensitive to physical activity changes in overweight/obese women. *International Journal of Obesity, 37*(12), 1597–1602.
Kirby, K. N., & MarakoviĆ, N. N. (1996). Delay-discounting probabilistic rewards: Rates decrease as amounts increase. *Psychonomic Bulletin & Review, 3*(1), 100–104.
Koffarnus, M. N., & Bickel, W. K. (2014). A 5-trial adjusting delay discounting task: Accurate discount rates in less than one minute. *Experimental and Clinical Psychopharmacology, 22*(3), 222.
Kotabe, H. P., & Hofmann, W. (2015). On integrating the components of self-control. *Perspectives on Psychological Science: A Journal of the Association for Psychological Science, 10*(5), 618–638.
Krebs, P., Prochaska, J. O., & Rossi, J. S. (2010). A meta-analysis of computer-tailored interventions for health behavior change. *Preventive Medicine, 51*(3–4), 214–221.
de Lauzon, B., Romon, M., Deschamps, V., Lafay, L., Borys, J.-M., Karlsson, J., ... Fleurbaix Laventie Ville Sante Study Group. (2004). The Three-Factor Eating Questionnaire-R18 is able to distinguish among different eating patterns in a general population. *The Journal of Nutrition, 134*(9), 2372–2380.
de Leeuw, J. R. (2015). jsPsych: A JavaScript library for creating behavioral experiments in a Web browser. *Behavior Research Methods, 47*(March 2014), 1–12.
Li, C., & Li, H. (2008). Network-constrained regularization and variable selection for analysis of genomic data. *Bioinformatics, 24*(9), 1175–1182.
Lu, C.-H., & Proctor, R. W. (1995). The influence of irrelevant location information on performance: A review of the simon and spatial Stroop effects. *Psychonomic Bulletin & Review, 2*(2), 174–207.
von Luxburg, U., Williamson, R. C., & Guyon, I. (2012). Clustering: Science or art? *JMLR Workshop and Conference Proceedings, 27*, 65–79.
Lynam, D. R., Smith, G. T., Whiteside, S. P., & Cyders, M. A. (2006). *The UPPS-P: Assessing five personality pathways to impulsive behavior.* West Lafayette, IN: Purdue University.
Mischel, W., Shoda, Y., & Rodriguez, M. (1989). Delay of gratification in children. *Science, 244*(4907), 933–938.
Miyake, A., & Friedman, N. P. (2012). The nature and organization of individual differences in executive functions. *Current Directions in Psychological Science, 21*(1), 8–14.
Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, a. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex "frontal lobe" tasks: A latent variable analysis. *Cognitive Psychology, 41*(1), 49–100.
Moffitt, T. E., Arseneault, L., Belsky, D., Dickson, N., Hancox, R. J., Harrington, H., ... Caspi, A. (2011). A gradient of childhood self-control predicts health, wealth, and public safety. *Proceedings of the National Academy of Sciences, 108*(7), 2693–2698.
Muchinsky, P. M. (1996). The correction for attenuation. *Educational and Psychological Measurement, 56*(1), 63–75.
Muraven, M., & Baumeister, R. F. (2000). Self-regulation and depletion of limited resources: Does self-control resemble a muscle? *Psychological Bulletin, 126*(2), 247–259.

Nee, D. E., Jonides, J., & Berman, M. G. (2007). Neural mechanisms of proactive interference-resolution. *NeuroImage, 38*(4), 740–751.
Nevill, A. M., Stewart, A. D., Olds, T., & Holder, R. (2006). Relationship between adiposity and body size reveals limitations of BMI. *American Journal of Physical Anthropology, 129*(1), 151–156.
Otto, A. R., Skatova, A., Madlon-Kay, S., & Daw, N. D. (2013). Cognitive control predicts use of model-based reinforcement learning. *Journal of Cognitive Neuroscience, 27*(2), 319–333.
Pashler, H. (1994). Dual-task interference in simple tasks: Data and theory. *Psychological Bulletin, 116*(2), 220.
Patrick, C. J., Curtin, J. J., & Tellegen, A. (2002). Development and validation of a brief form of the multidimensional personality questionnaire. *Psychological Assessment, 14*(2), 150–163.
Patton, J. H., Stanford, M. S., & Barratt, E. S. (1995). Factor structure of the Barratt impulsiveness scale. *Journal of Clinical Psychology, 51*(6), 768–774.
Pleskac, T. J. (2008). Decision making and learning while taking sequential risks. *Journal of Experimental Psychology. Learning, Memory, and Cognition, 34*(1), 167–185.
Poldrack, R. A., Kittur, A., Kalar, D., Miller, E., Seppa, C., Gil, Y., ... Bilder (2011). The cognitive atlas : Toward a knowledge foundation for cognitive neuroscience. *Frontiers in Neuroinformatics, 5*, 17.
Poldrack, R. A., & Yarkoni, T. (2016). From brain maps to cognitive Ontologies: Informatics and the search for mental structure. *Annual Review of Psychology, 67*, 587–612.
Primi, C., Morsanyi, K., Chiesi, F., Donati, M. A., & Hamilton, J. (2016). The development and testing of a new version of the cognitive reflection test applying item response theory (IRT). *Journal of Behavioral Decision Making, 29*(5), 453–469.
Ratcliff, R., & Childers, R. (2015). Individual differences and fitting methods for the two-choice diffusion model of decision making. *Decisions, 2*(4) Advance online publication.
Ratcliff, R., Smith, P. L., Brown, S. D., & McKoon, G. (2016). Diffusion decision Model: Current issues and history. *Trends in Cognitive Sciences, 20*(4), 260–281.
Raven, J., & Raven, J. (2003). Raven progressive matrices. In R. S. McCallum (Ed.). *Handbook of nonverbal assessment* (pp. 223–240). New York: Kluwer Academic/ Plenum Publishers.
van Ravenzwaaij, D., & Oberauer, K. (2009). How to use the diffusion model: Parameter recovery of three methods: EZ, fast-dm, and DMAT. *Journal of Mathematical Psychology, 53*(6), 463–473.
Revelle, W. (2017). *psych: Procedures for personality and psychological research.* Evanston, Illinois, USA: Northwestern University. https://CRAN.R-project.org/package=psych Version = 1.7.5.
de Ridder, D., & de Wit, J. (2006). *Self-regulation in health behavior.* John Wiley & Sons.
Roth, R. M., Isquith, P. K., & Gioia, G. A. (2005). *BRIEF-A: Behavior rating inventory of executive function–adult version : Professional manual.* Psychological Assessment Resources.
Schneider, D. W., & Logan, G. D. (2011). Task-switching performance with 1:1 and 2:1 cue-task mappings: Not so different after all. *Journal of Experimental Psychology. Learning, Memory, and Cognition, 37*(2), 405–415.
Shallice, T. (1982). Specific impairments of planning. *Philosophical transactions of the royal society of London, Biology, 298*, 199–209.
Sharma, L., Markon, K. E., & Clark, L. A. (2013). Toward a theory of distinct types of "impulsive" behaviors: A meta-analysis of self-report and behavioral measures. *Psychological Bulletin, 140*(2), 374–408.
Shiffman, S., Stone, A. A., & Hufford, M. R. (2008). Ecological momentary assessment. *Annual Review of Clinical Psychology, 4*, 1–32.
Sochat, V. V., Eisenberg, I. W., Enkavi, A. Z., Li, J., Bissett, P. G., & Poldrack, R. A. (2016). The experiment Factory: Standardizing behavioral experiments. *Frontiers in Psychology, 7*, 610.
Stahl, C., Voss, A., Schmitz, F., Nuszbaum, M., Tüscher, O., Lieb, K., et al. (2014). Behavioral components of impulsivity. Journal of experimental psychology. *General, 143*(2), 850–886.
Stanford, M. S., Mathias, C. W., Dougherty, D. M., Lake, S. L., Anderson, N. E., & Patton, J. H. (2009). Fifty years of the barratt impulsiveness Scale: An update and review. *Personality and Individual Differences, 47*(5), 385–395.
Stekhoven, D. J., & Buhlmann, P. (2012). MissForest–non-parametric missing value imputation for mixed-type data. *Bioinformatics, 28*(1), 112–118.
Toplak, M. E., West, R. F., & Stanovich, K. E. (2014). Assessing miserly information processing: An expansion of the Cognitive Reflection Test. *Thinking & Reasoning, 20*(2), 147–168.
Vohs, K. D., & Baumeister, R. F. (2016). *Handbook of self-regulation: research, theory, and applications.* Guilford Publications.
Wagenmakers, E.-J., van der Maas, H. L. J., & Grasman, R. P. P. P. (2007). An EZ-diffusion model for response time and accuracy. *Psychonomic Bulletin & Review, 14*(1), 3–22.
Whitley, D. (1994). A genetic algorithm tutorial. *Statistics and Computing, 4*(2), 65–85.
Wickelgren, W. A. (1977). Speed-accuracy tradeoff and information processing dynamics. *Acta Psychologica, 41*(1), 67–85.
Wiecki, T. V., Sofer, I., & Frank, M. J. (2013). HDDM: Hierarchical bayesian estimation of the drift-diffusion model in Python. *Frontiers in Neuroinformatics, 7*(August), 14.
Wilson, R. C., & Niv, Y. (2012). Inferring relevance in a changing world. *Frontiers in Human Neuroscience, 5*(January), 1–14.
Woods, D. L., Kishiyama, M. M., Yund, E. W., Herron, T. J., Edwards, B., Poliva, O., et al. (2011). Improving digit span assessment of short-term verbal memory. *Journal of Clinical and Experimental Neuropsychology, 33*(1), 101–111.
Yntema, D. B. (1963). Keeping track of several things at once. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 5*(1), 7–17.
Zimbardo, P. G., & Boyd, J. N. (2015). Putting time in perspective: A valid, reliable individual-differences metric. *Time perspective theory; review, research and application* (pp. 17–55). Springer.
Zuckerman, M. (2007). The sensation seeking scale V (SSS-V): Still reliable and valid. *Personality and Individual Differences, 43*(5), 1303–1305.