Mobile Data Collection of Cognitive-Behavioral Tasks in Substance Use Disorders: Where Are We Now?

Hilmar G. Zech[a]   Markus Reichert[b, c, d]   Ulrich W. Ebner-Priemer[c, d]   Heike Tost[d]
Michael A. Rapp[e]   Andreas Heinz[f]   Raymond J. Dolan[g, h]   Michael N. Smolka[a]
Lorenz Deserno[a, i, j]

[a]Department of Psychiatry, Technische Universität Dresden, Dresden, Germany; [b]Department of eHealth and Sports Analytics, Faculty of Sport Science, Ruhr-Universität Bochum (RUB), Bochum, Germany; [c]Institute of Sports and Sports Science, Karlsruhe Institute of Technology (KIT), Mental mHealth Lab, Karlsruhe, Germany; [d]Central Institute of Mental Health, Department of Psychiatry and Psychotherapy, Medical Faculty Mannheim, Heidelberg University, Mannheim, Germany; [e]Department for Social and Preventive Medicine, University of Potsdam, Potsdam, Germany; [f]Department of Psychiatry and Psychotherapy, Charité Campus Mitte, Charité – Universitätsmedizin Berlin, Corporate member of Freie Universität Berlin and Humboldt-Universität zu Berlin, Berlin, Germany; [g]Max Planck UCL Centre for Computational Psychiatry and Ageing Research, University College London, London, UK; [h]Wellcome Centre for Human Neuroimaging, University College London, London, UK; [i]Department of Child and Adolescent Psychiatry, Psychotherapy and Psychosomatics, University of Würzburg, Würzburg, Germany; [j]Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany

Abstract

Introduction: Over the last decades, our understanding of the cognitive, motivational, and neural processes involved in addictive behavior has increased enormously. A plethora of laboratory-based and cross-sectional studies has linked cognitive-behavioral measures to between-subject differences in drinking behavior. However, such laboratory-based studies inevitably suffer from small sample sizes and the inability to link temporal fluctuations in task measures to fluctuations in real-life substance use. To overcome these problems, several existing behavioral tasks have been transferred to smartphones to allow studying cognition in the field. *Method:* In this narrative review, we first summarize studies that used existing behavioral tasks in the laboratory and self-reports of substance use with ecological momentary assessment (EMA) in the field. Next, we review studies on psychometric properties of smartphone-based behavioral tasks. Finally, we review studies that used both smartphone-based tasks and self-reports with EMA in the field. *Results:* Overall, studies were scarce and heterogenous both in tasks and in study outcomes. Nevertheless, existing findings are promising and point toward several methodological recommendations: concerning psychometrics, studies show that – although more systematic studies are necessary – task validity and reliability can be improved, for example, by analyzing several measurement sessions at once rather than analyzing sessions separately. Studies that use tasks in the field, more-

Correspondence to:
Hilmar G. Zech, hilmar.zech@tu-dresden.de

over, show that power can be improved by choosing sampling schemes that combine time-based with event-based sampling, rather than relying on time-based sampling alone. Increasing sampling frequency can further increase power. However, as this also increases the burden to participants, more research is necessary to determine the ideal sampling frequency for each task. ***Conclusion:*** Although more research is necessary to systematically study both the psychometrics of smartphone-based tasks and the frequency at which task measures fluctuate, existing studies are promising and reveal important methodological recommendations useful for researchers interested in implementing behavioral tasks in EMA studies.

Introduction

Substance use, including alcohol, tobacco, and illicit drugs, is one of the leading risk factors for death and disability worldwide [1]. One of the difficulties in understanding and treating addiction is that multiple processes are likely to be at work at different stages of addiction or even in parallel [2]. Further, many of the processes underlying substance use are not directly accessible to conscious awareness [3, 4], thus cannot easily be assessed with interviews or self-reports. Instead, behavioral tasks, in combination with self-reports, have been used to tap into the processes associated with addiction implicitly. Such tasks have already been used to link substance use, for example, to reduced cognitive control [5], attentional and behavioral biases toward drugs of abuse [6, 7], and biased implicit attitudes to drug stimuli [8]. Moreover, behavioral measures have been linked to neural markers and versions of behavioral tasks have been used to create promising interventions targeting substance use [9]. However, despite the considerable progress in neurocognitive addiction research, the impact on clinical diagnosis or treatment has been rather limited.

Today, most task-based studies in addiction research have been limited to cross-sectional laboratory studies [10]. These types of studies have several limitations: first, addictive behaviors do not commonly happen inside the laboratory, thus, studies have to rely on retrospective self-reports to link task measures to addictive behaviors. Such retrospective measures can suffer from recall bias and heuristics causing inaccurate or biased findings [11–13]. Second, many substance use-related processes undergo significant temporal fluctuations, and substance use behaviors are influenced considerably by contexts such as mood or social contexts, which are difficult, if not impossible, to capture in the laboratory [13]. Cross-sectional laboratory studies usually cannot tap into such temporal or context-dependent dynamics of behavioral variables, making it almost impossible to understand whether temporal changes in substance use are driven by changes in these variables or to understand how these variables are influenced by substance-related contexts. Finally, it is not known whether changes in mechanisms identified in the laboratory also play a role in people's real lives – a question that cannot be answered in artificial laboratory environments.

To overcome these problems, several researchers have propagated the use of ecological momentary assessments (EMAs) (Fig. 1) [13–15]. In EMA, measures are taken repeatedly in people's natural environments with ideally no or as much as possible reduced latency, thereby decreasing recall bias and increasing ecological validity, thus allowing researchers to study temporal fluctuations and context-dependent effects [16]. These features of EMA are especially useful in addiction research and, indeed, an increasing number of self-report-based EMA studies have already been conducted in addiction research and have yielded valuable insights (for an overview see Shiffman [13]). Most reviews of EMA-based studies distinguish three groups of EMA: Diary methods, continuous physiological monitoring (for example, heart rate), and activity tracking (for example, physical activity and location tracking) [13, 15]. A possible fourth group, EMA studies based on behavioral tasks, has so far received little attention. This is likely due to the fact that classical behavioral tasks mostly depend on stationary equipment (for example, desktop computers), which cannot easily be deployed in the field. Modern smartphones can overcome this shortcoming as they have specific benefits for implementing task-based EMA studies (for a summary, see Miller [17]): For example, their high computational power and ability to display rich stimuli and record rich behavioral responses make precise measurements of, for example, reaction times possible. Moreover, their connectivity and ubiquity – as they are already carried by most potential participants virtually 24 h/day – make them ideal for EMA studies in the field. Consequently, there are already notable examples of assessment of behavioral tasks using dedicated smartphone apps to examine large-scale samples via app stores [18, 19]. Yet, only a few behavioral task-based EMA studies have been conducted in the context of addiction research so far. To outline the potential of using smartphones to run behavioral tasks in EMA studies, the aim
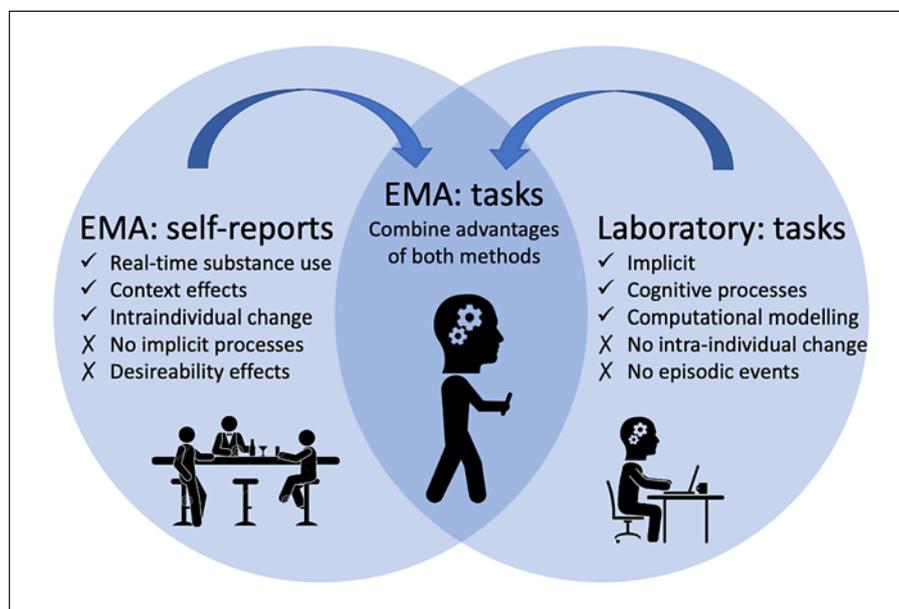
Fig. 1. Advantages of using behavioral tasks during EMAs of substance use.

of the current review is to give an overview of these task-based EMA studies and discuss arising design limitations and methodological issues.

Mobile Data Collection of Behavioral Tasks in Substance Use Disorders: Where Are We Now?

In this narrative review[1], we categorize the available studies into three groups: (1) studies that used behavioral tasks in the laboratory and self-reports with EMA in the field; (2) studies that assessed the psychometric properties of mobile behavioral tasks; and (3) studies that used both tasks and self-reports with EMA in the field.

While reviewing these studies, we focus on several methodological decisions, such as the applied sampling rate (how often data are collected), sampling scheme (whether sampling intervals are regular, random, or event-based), and level analysis (participant level, session level, trial level). The chosen sampling rate and scheme as well as level of analysis can have significant effects on study outcomes. For example, Shiffman et al. [20] showed

that although daily negative affect did not predict relapse, hourly change in affect (on the same day of the relapse) did. The sampling rate is also important. Time-based schedules (either at specific times or randomly scheduled) allow researchers to get a representative sample of participants' behavior. Yet, when sampling rate is low, time-based schedules might miss rare but potentially impactful events, such as binge drinking and relapse, and assessment of these may still suffer from significant recall bias [13, 15]. Event-based schedules, in which participants are instructed to initiate measurements when a certain event occurs can overcome this problem. They should be complemented with time-based schedules, though, to determine whether effects are specific to events [13]. Finally, continuous measurements can be used for data that can be collected passively (motion, location), or physiological measures and interactive schedules could automatically trigger task measures when certain physiological or context criteria are met. However, foreshadowing insights from our literature review, these potentials of EMA designs in combination with behavioral assessments are not yet exploited sufficiently in the context of addiction research.

Section 1: Tasks in the Laboratory and Self-Reports of Substance Use with EMA in the Field

We identified 8 studies (Table 1) that have used tasks in the laboratory and associated their readouts with self-reported EMA data on substance use collected in the field.

---

[1] Although based on PRISMA guidelines this is not sufficient for a systematic review, we conducted a literature search on PubMed with the following search terms (alcohol OR addiction OR "use disorder" OR "binge drinking") AND ("ecological momentary" OR "smartphone" OR "experience sampling" OR "mobile phone" OR "ambulatory assessment"). After reviewing titles of all 637 search results and abstracts of 236 articles, we included 18 articles in this review.

Zech/Reichert/Ebner-Priemer/Tost/Rapp/
Heinz/Dolan/Smolka/Deserno

Table 1. Overview of studies that use tasks in the laboratory and self-reports in EMA

Study	Sample (analyzed)	Task(s)	EMA self-report	EMA period	Sampling rate	Findings
Mereish et al. [21][a]	55 cannabis users	Startle response	Cannabis craving, cue exposure, cannabis use	7 days	3-hourly	Startle responses predict craving
Miranda et al. [22][a]	85 cannabis users	WM	Cannabis craving, cue exposure, cannabis use, stress	7 days	3-hourly	WM reduces the effect of stress on craving
Treloar Padovano et al. [23][a]	86 frequent drinkers	WM	Alcohol craving, cue exposure, alcohol use, stress	7 days	3-hourly	WM reduces the effect of stress on craving
Begh et al. [25]	100 smokers	Stroop, visual probe	Cigarette craving, cue exposure, attention to smoking	7 weeks	3 per day	–
Groefsema et al. [26]	192 social drinkers	SR-compatibility, visual probe	Alcohol use	5 weeks	3 per week	–
Hendriks et al. [28]	702 students	Odd one out, spatial span, spatial, rotation, digit span, paired association, tree task	Alcohol use	4 weeks[b]	3 per week	–
Shiffman et al. [24]	190 daily smokers	Cue reactivity	Smoking cue/prohibition exposure, affect, cigarette consumption	6 weeks	5 per day	–
Snelleman et al. [27]	50 alcohol-dependent	Stroop, SR-compatibility	Relapse	3 months	1 per month	–

This table shows review studies that used tasks in the laboratory and self-reports in EMA. SR, stimulus-response. [a] These studies all used the subject pool collected in a study by Miranda et al. [22] (2016). [b] Both tasks and the EMA period were completed twice with a 1-year interval.

For example, Mereish et al. [21] measured appetitive startle responses to cannabis pictures in the laboratory using a startle response task. Next, participants reported cannabis cue exposure and craving, in the field. Using this method, Mereish et al. [21] found that reduced appetitive startle responses were associated with a reduced influence of cue exposure on craving. In the same sample, Miranda et al. [22] measured working memory (WM) in the laboratory and found that increased WM performance/capacity was associated with a reduced influence of stress on cannabis craving in the field. Using the same WM task in the lab, Treloar Padovano et al. [23] found that WM also reduces the effect of stress on alcohol craving in drinkers (but only in males). Together these 3 studies show how behavioral between-participant variables measured at baseline in the laboratory can moderate relationships between self-reported variables measured longitudinally in the field.

The subsequently described studies employed similar designs, however, did not detect significant relationships between laboratory-based behavioral measures and EMA-based self-reports (Table 1). Shiffman et al. [24] measured cue reactivity to smoking cues in the laboratory using self-reports. Next, participants reported exposure to smoking cues, positive and negative affect, exposure to smoking prohibitions, and cigarette consumption in the field. However, Shiffman et al. [24] found no relationship between baseline in-laboratory and EMA variables. Another study by Begh et al. [25] used a Stroop task and a visual probe task to measure attentional biases in the laboratory. Next, participants reported exposure to smoking cues, attention to smoking, and craving for cigarettes in the field. Yet, Begh et al. [25] found no association between attentional biases and neither of the EMA variables. Groefsema et al. [26] used a stimulus-response-compatibility task and a visual probe task in the laboratory to measure approach and attentional biases. Participants reported their alcohol use in the field, but no association between biases and consumption was found. Snelleman et al. [27] used a stimulus-response-compatibility task and a Stroop task in the laboratory to measure approach and attentional biases. They found no relation-

Table 2. Overview of psychometric studies

Study	Sample (analyzed)	Task(s)	Platform(s)	EMA period	Sampling rate	Sampling scheme	Findings
Bouvard et al. [36]	34 participants with substance disorder, 27 healthy control	Stroop, verbal fluency	Android	1 week	4/day	Time-based (every 3 h)	Moderate construct validity, correlations between phone and computer tasks (rs >0.42)
Pal et al. [38]	16 methamphetamine users, 20 healthy control	N-back, stop-signal, and verbal stroop	iOS	2 weeks	2/day	Time-based (participant specified)	Task measures did not differ between platforms (phone vs. computer)
Smith et al. [33]	19 healthy participants	Balloon analog risk	iOS/Android	2 weeks	>1/day	Time-based and event-based	Adequate compliance (78%); low reactivity and good usability
Spanakis et al. [40]	120 social drinkers	Stroop	Android	–	–		Internal consistencies high enough for basic research (α >0.70)
Jones et al. [39]	100 heavy drinkers	Stop-signal	Android	2 weeks	2/day	Random	High test-retest reliability based on average scores (α = 0.96)
Sliwinski et al. [37]		Symbol search task, dot memory task, N-back task	Android	2 weeks	5/day	Quasi-random every 2–3 h	High test-retest reliability based on average scores (ICCs >0.97) but low based on individual scores (ICCs 0.41–0.53); good construct validity

This table shows reviewed psychometric studies.

ship with relapse reported in the months after this measurement. Hendriks et al. [28] used six cognitive performance tasks (see Table 1) in the laboratory. Next, participants reported alcohol consumption in the field, but no association between the tasks and EMA measures was found.

Interim Conclusion Section 1

As summarized in Table 1, some of the studies were performed in smaller-to-moderate sample sizes (Ns < 100), observation periods were in the range of weeks, and there was strong heterogeneity in the cognitive-behavioral constructs studied. This substantially limits the possibility to draw firm conclusion both with respect to implications of the positive and negative findings reported. While the total number of studies was low and heterogenous (for example, most likely unsuitable for meta-analysis), it should be noted that it was the highest number of studies identified as compared to Sections 2 and 3. This indicates the still early state of mobile data collection of behavioral tasks. More generally, the discussed studies rely on the assumption that constructs underlying task-based measures are temporally stable – at least stable enough to explain fluctuation in EMA self-reports. This assumption might not be warranted. It has, for example, been hypothesized that self-control, as captured also by behavioral tasks, fluctuates throughout the day [29, 30].

From a statistical viewpoint, it has been noted that many tasks have low test-retest reliability (ICCs <0.5), which – among other implications – indicates that they might not be temporally stable (or they might be generally unreliable) [31]. However, low – test-retest reliability puts an "upper bound" on the maximum correlation that can be detected [32]. This may be one important explanation to consider with regard to the weak or null effects reported. Moreover, studies that measure behavioral measures cross-sectionally in the laboratory cannot assess the temporal dependency between behavioral and self-report measures (although they can be associated with aggregated measures of change). Studies that deploy behavioral tasks longitudinally on mobile devices in the field can (partially) overcome these limitations. These studies require versions of behavioral tasks that can be completed in the field (for example, on smartphones). We will next review studies that compare these smartphone-based versions to laboratory-based tasks, before giving an overview of existing studies using tasks in the field.

Section 2: Feasibility and Psychometric Substance Use Studies

In this section (see Table 2 for an overview), we first briefly report on 1 study that did not examine psychometric properties, in the strict sense, but rather assessed compliance or feasibility – two aspects of tasks that are espe-

cially important in EMA research, where high compliance is of special interest. In a pure feasibility study, Smith et al. [33] developed a smartphone version of the balloon analog risk task. They found that compliance was adequate (78% of expected responses completed). Reactivity (that is, the extent to which behavior is affected by the assessment) was reported as low and usability as high. To increase compliance, Smith et al. [33] used a mixed sampling scheme in which participants initiated the app at the beginning of a drinking session (event-based) and were consequently asked every hour (time-based) how much alcohol they consumed (until they manually ended the drinking session).

Psychometric studies have recently started to gain more attention in neurocognitive and neuropsychiatric research [31, 34, 35]. This is because sufficient psychometric properties (that is, validity and reliability) of any measurement variable are necessary to study interindividual differences (the main target of research in psychiatry and addiction) in a clinically meaningful manner.

With respect to the validity of mobile data collection of behavioral tasks (Table 2), the behavioral readouts should be correlated with laboratory-based assessments of the same task. Bouvard et al. [36] developed smartphone-based versions of the verbal Stroop and verbal fluency tasks, in which participants recorded verbal responses on the smartphone. They found that data from these tasks collected in the field correlated with data from laboratory-based validation tasks. It should be noted that the correlation was, on average, lower for patients with substance disorder (Stroop: $r = 0.70$; Verbal fluency: $r = 0.42$) than for healthy control participants (Stroop: $r = 0.90$; Verbal fluency: $r = 0.68$). Sliwinski et al. [37] developed smartphone-based versions of the symbol search task to measure perceptual speed and the dot memory task as well as the n-back task to measure WM. Healthy participants completed the tasks in a 14-day EMA period in the field. They tested their tasks' construct validity by asking participants to complete several related behavioral tasks in the laboratory. They report good construct validity as tasks correlated significantly with in-laboratory tasks (rs ranging from 0.39 to 0.74). Pal et al. [38] compared smartphone and computer-based versions of the n-back, stop-signal, and verbal Stroop task. It should, however, be noted that instead of correlating these tasks with each other, they only tested for differences between smartphone and smartphone-based tasks, which is not how construct validity is properly assessed. For the stop-signal task, they did not find differences between task measures from different platforms (due to programing

errors data from the n-back and Stroop task could not be compared).

There has been a particular focus on test-retest reliability (hereafter referred to simply as reliability) – or the temporal stability of a measurement across at least two test sessions. This is because the reliability of two variables studied (for example, behavior in a task and craving from a questionnaire) essentially puts an upper bound on the maximum correlation that can be detected between variables [32]. In plane words, if reliability is low, we cannot detect meaningful correlations between brain-behavior variables and clinical symptoms. In the context of EMA, it is important to distinguish within- and between-person reliability [37]. Between-person reliability is important when correlating task measures with between-person (that is, trait) measures. Within-person reliability, on the other hand, is important when correlating task measures with within-person (that is, state) measures. For example, in a notable study on test-retest reliability (Table 2), Sliwinski et al. (see above [37]) found that tasks had excellent test-retest reliability (ICCs >0.97) when relying on average measures from multiple measurement sessions (between-person), but poor to moderate reliability when basing measures on a single measurement session (within-person; ICCs ranging from 0.41 to 0.53). Jones et al. [39] developed a smartphone-based version of the stop-signal reaction time (SSRT) task that participants completed two times per day over two weeks. Similar to Sliwinski, they report excellent test-retest reliability when calculating scores based on averages (Cronbach's α = 0.96). On the one hand, this finding indicates that task measures might undergo significant state fluctuations, which need to be better understood in future research. On the other hand, it shows that – until fluctuations can be predicted – increasing the number of measurement sessions can increase the reliability of task measures to acceptable levels.

Another aspect of reliability – internal consistency – describes how consistently a task measures its construct within one measurement session. In the context of EMA, this characteristic is especially important, as low test-retest reliability combined with high internal consistency can indicate that a task can measure a construct with little measurement error, but that the underlying construct is subject to potentially meaningful change over time [31]. In this regard (Table 2), Spanakis et al. [40] developed two smartphone versions of the color Stroop task (one with words and one with pictures of alcohol). Half of the participants completed these mobile versions at home and half-completed computer-based versions in the labora-

Table 3. Overview of studies using tasks and self-reports in EMA

Study	Sample (analyzed)	EMA Task(s)	Platform(s)	EMA self-report	EMA period	Sampling rate	Sampling scheme	Analysis	Findings
Jones et al. [39]	100 heavy drinkers	Stop-signal	Android	Alcohol consumption	2 weeks	2/day	Random	Mixed models, participant-mean centered	Decrease of inhibition predicts consumption ($\beta = 0.007$)
Emery et al. [43]	92 moderately drinking college students	Stroop	iOS/ Android	Alcohol consumption	4 weeks	9/day	Time-based (random within 2-h blocks)	Multilevel structural equation models, aggregated into day and nighttime	Nighttime attentional biases predict consumption (IRR = 1.11)
Marhe et al. [47]	68 heroin-dependent inpatients	Stroop, implicit association	PDA	Relapse	1 week	>4/day	Time-based (random) and event-based	Mixed models	Behavioral variables predict relapse at times of temptation ($b_{Stroop} = 0.0053$; $b_{IAT} = 0.00055$)
Suffoletto et al. [48]	219 risky drinkers	Stroop, approach-avoidance	N.D.	Binge drinking	14 weeks	1/week	Time-based (Thursday 3 p.m.)	Mixed models, participant-mean centered	–
MacLean et al. [49]	20 heavy drinking smokers	Dot-probe	PDA	Alcohol consumption, smoking	1 week	4/day	Random (within four equal blocks of waking time)	Mixed models	–

This table shows reviewed studies that used tasks as well as self-reports in EMA. IAT, implicit association test.

tory. The internal consistency of smartphone-based Stroop tasks was high enough for early basic research but not high enough for applied settings such as clinical diagnostics ($\alpha = 0.70$ for alcohol words and $\alpha = 0.74$ for alcohol pictures; based on qualitative interpretations by Nunnally et al. [41]; for a discussion of those interpretations see Lance et al. [42]). It is noteworthy that the smartphone task's internal consistency was higher than that of computer-based Stroop tasks ($\alpha = 0.49$ for alcohol words and $\alpha = 0.58$ for alcohol pictures). Emery et al. [43] also developed a smartphone-based version of the Stroop task and tested participants over a period of four weeks (for a full description of the study see Section 3). Unlike Spanakis et al. [40], they found that task's internal consistency was poor ($r_{sb} = 0.26$). It should be noted that Emery et al. [43] implementation of the Stroop task differed from that of Spanakis' et al. [40] version, as participants did not have to respond to colors as in the classical Stroop task, but rather had to count the number of words displayed on the screen.

Interim Conclusion Section 2

Studies on the psychometric properties of task-based EMA studies in addiction research are overall scarce.

The 6 studies reviewed were heterogeneous. One study reported on general compliance and feasibility readouts to be overall good. Construct validity of smartphone-based tasks was examined in 1 study and reported to be good with smartphone-based tasks showing moderate to strong correlations with computer-based counterparts. Overall, the level of reporting on psychometric properties was inconsistent and most studies focused on single tasks which differed from those tested in other studies. A notable study in healthy participants that applied a set of tasks indicated that test-retest reliability could be improved but only when relying on several rather than single measurements. The way data are either aggregated or modeled has been identified as one critical factor in the context of reliability [18, 44–46]. Therefore, we refer for this aspect to the more general discussion. Studies on internal consistency varied greatly, with 1 study reporting low internal consistencies and one reporting internal consistency high enough for basic research. In sum, while it may be possible to reach sufficient psychometric properties with task-based EMA studies, it has to be emphasized that this needs to be studied more systematically.

Zech/Reichert/Ebner-Priemer/Tost/Rapp/
Heinz/Dolan/Smolka/Deserno

Section 3: Tasks and Self-Reports of Substance Use with EMA in the Field

Out of 5 studies (Table 3), 3 studies reported positive associations between task measures and measures of substance use. Jones et al. [39] tested participants with a smartphone-based version of the SSRT task to measure inhibition and measured alcohol consumption based on self-reports (measured retrospectively for the preceding day). Although they did not find a direct association between inhibition and alcohol consumption, they found that a decreased inhibition on a given day is associated with increased alcohol consumption on that day. This finding shows that the way in which EMA data is aggregated can have a substantial impact on observed effects. Emery et al. [43] tested participants two times per day with a smartphone-based Stroop task and asked them to report their overall alcohol consumption on the previous day. They found that nighttime but not daytime attentional biases to alcohol correlated with drinking behavior on the same day. Marhe et al. [47] developed PDA-based versions of the Stroop task and the implicit association test. They collected behavioral measures both at random times and when participants indicated to be tempted by drug stimuli. They found that at random times, behavioral variables did not predict relapse; however, they did predict relapse at times of temptation. This study emphasizes how different sampling schemes (time-based vs. event-based) can reveal different effects.

Suffoletto et al. [48] used smartphone versions of the Stroop task and the approach-avoidance task. They found no relationship to addictive behavior as neither attentional biases nor approach biases correlated with neither binge drinking. MacLean et al. [49] developed a PDA-based version of the dot-probe task and found no association between attentional biases with neither smoking nor drinking behavior (note, however, that the sample size in this study was very low).

Interim Conclusion Section 3

Task-based EMA studies in the context of substance use are also scarce. However, three out of 5 studies indicated promising findings such that measures of cognitive control (SSRT & Stroop) related to changes in measures of addictive behavior. These studies also demonstrate how the level of analysis (how data are aggregated) as well as sampling rate and sampling scheme (when task measures are collected) can have a significant influence on results. Specifically, higher sampling rates and combinations between event- and time-based sampling schemes seem to be more powerful than lower sampling rates and

simple time-based sampling schemes. Interestingly, the two null findings were both using measures of attention. On the one hand, this could indicate that the psychometric criteria of these tasks are generally low (note that the authors of the 2 studies reviewed here do not report these criteria, but some studies indeed report low reliability for laboratory-based versions of the tasks [50]). On the other hand, it could indicate that either task-based measures of attention are not as important in moderating change in self-reported substance use or putatively reflecting that these tasks may be less well applicable in the field. This conclusion resonates with data from an attention task in very a large app store-based sample [18]. It should be noted that most of the studies reviewed in this section did not focus on patient samples. While focusing on healthy participants is advantageous in early stages of research as it allows for larger samples and more convenient data collection, future studies should replicate findings in patient samples to see if they generalize to clinical populations.

Discussion

In this narrative review, we summarized studies in the context of substance use relying on cognitive-behavioral tasks in the context of mobile data collection or EMAs. First, we reviewed 8 studies that used behavioral tasks in the laboratory and correlated behavioral measures to self-report measures of substance use in the field. These studies overall showed weak or no links between behavioral measures and measures of substance use. Second, we reviewed 6 studies in the section on psychometric properties of mobile behavioral tasks. While studies on construct validity showed that mobile tasks mostly reflected measurements taken by computer-based tasks, studies on internal consistency yielded mixed results, ranging from poor to excellent internal consistency. There was only one study on test-retest reliability, which indicated that reliability was low when calculating task variables based on single measurement (within-person reliability) sessions but high when task variables were based on averages of several measurement sessions (between-person reliability). In sum, more studies on psychometrics are needed – especially studies that compare different tasks in the same sample. Third, we reviewed studies that used both mobile behavioral tasks and self-report measures in the field. Although such studies are scarce, results were overall promising by indicating that especially tasks measuring cognitive control can capture changes in measures of substance use. These studies moreover indicate that ef-

fects are highly dependent on sampling schemes and sampling rates – with more effects reported for event-based sampling schemes and high sampling rates.

The reviewed studies revealed several methodological issues that researchers should consider when using behavioral tasks in the context of EMA. First, there were only few systematic studies of psychometric properties available. A single study on test-retest reliability showed that mobile behavioral tasks' reliability was low at least when estimated based on single sessions. This mirrors the laboratory-based finding that the reliability of computer-based behavioral tasks tends to be overall low [31]. Mobile tasks might have an advantage over computer-based tasks because they allow researchers more easily to collect several assessments and aggregate data which significantly increased reliability [37]. This procedure is intuitive when considering that more data points will, at least to some degree, improve stability of a measurement [18,44, 46, 51]. However, this limits clinical application as well as application on the smartphone because for patients and studies in the field, tasks should plausibly be rather on the shorter end of duration in order to decrease burden to participants. Related to this, it was also pointed out that avoiding aggregation and to model data as detailed as possible, thus, on the trial-by-trial level (the "highest possible resolution") is advantageous for the assessment of reliability [45, 46]. This may be because more comprehensive modeling of single trials obtained from multiple sessions captures the data, in particular its within- and between-subjects variance sources, in a manner as it was collected. Such analytic approaches, described as hierarchical or generative modeling as well [45], are also promoted by researchers using computational models of behaviors, such as reinforcement learning [52], which describe trial-by-trial dynamics of learning and choice. As part of our new collaborative research center [2], we will soon have smartphone-based behavioral data available from multiple tasks to address some of these questions. We have also preregistered-related research questions [53].

Second, another methodological issue researchers should be aware of is that the choice of sampling rates and sampling schemes can have a substantial impact on detecting certain findings or not [47]. Choosing a sampling rate and sampling scheme appropriate to a given research question requires knowledge about the expected rate of change in behavioral variables. Given that few studies exist today that use behavioral tasks during EMA, this knowledge is often not available and still needs to be established. On the one hand, future studies could maximize sampling rates to capture potentially fast-changing behavioral variables and to avoid missing potentially impactful events (for example, binge drinking). On the other hand, it should also be considered that increased sampling rates also increase the burden to participants – which might in turn decrease compliance – and study costs. This tradeoff can potentially be eased by using adequate sampling schemes – such as combinations of time- and event-based schedules [33] or by making tasks easier to complete. One way to achieve this is decreasing task length, which however might again decrease within-person reliability. In this regard, Gamification elements can also make tasks more engaging [19], thereby counteract the burden caused by increased sampling rates, and hopefully also decrease measurement error – a plausible assumption that has not yet been tested specifically. As already pointed in the context of reliability, another challenge is that researchers have to carefully chose the level of analysis of EMA data. Both the period of time over which data are aggregated as well as the choice of aggregation function can significantly influence findings (for example, see Jones et al. [39] for more details).

In this review, we primarily focused on studies that used behavioral tasks in EMA studies of substance use. A limitation in this regard is that behavioral tasks that have already been used in EMA studies but have not yet been applied in the domain of substance use. For example, in the category of studies that used behavioral tasks in the laboratory and correlated behavioral measures to self-report measures in the field (see Section 2), several studies outside the domain of substance use have shown associations between task measures and self-reports of self-control failure [54–59]. However, associations in these studies have been overall weak, indicating that they might suffer from the same reliability and state-dependency issues discussed in the current review. In the category of studies that used both mobile behavioral tasks and self-report measures in the field (Section 3), several additional tasks have already been developed and tested outside the domain of substance use. In a recent review, McKinney et al. [60] gave an overview of 18 such tasks, including a smartphone-based Flanker task developed by Kennedy et al. [61], which could also be of interest in substance use research. Although most of these tasks measure similar constructs (for example, cognitive control, attention, WM) as those reviewed here, they might have better or worse psychometric qualities, which need to be established in future studies. Another example is a smartphone-based version of the approach-avoidance tasks, recently developed by Zech et al. [62], which in addition to classically measured reaction times uses the phone's built-in sensors to also de-

Zech/Reichert/Ebner-Priemer/Tost/Rapp/
Heinz/Dolan/Smolka/Deserno

tect response force – a measure frequently used in animal research on motivation [63]. Future studies could implement these tasks in the domain of substance use.

Another limitation is that we did not focus on studies that use physiological and activity (physical activity and location) measures in EMA. Rare examples in addiction research are, Jung et al. [64], who tested a smartphone-based sensor to measure saliva alcohol in the field, and Suffoletto et al. [65], who used smartphone-based gait analysis to measure intoxication. Sensor-based studies could be used to complement task-based studies in the future. To connect behavioral measures to physiological sensor data, computational models of reinforcement learning might be useful, which can capture the processes and mechanisms at work and are biologically plausible to inform the analysis of physiological data [66]. For example, Eldar et al. [67] used a smartphone-based learning task in combination with reinforcement learning models and wearable sensors to track healthy participants' mood, reward sensitivity, heart rate, and EEG over a period of 1 week. The decodability of reward sensitivity in physiological signals could predict subsequent fast as well as slow mood fluctuations in healthy participants. As dysregulated reward sensitivity has also been linked to addiction [68] and mood is coupled to craving [69–73] similar intense sampling methods that combine EMA-based tasks, computational models, and wearable sensors could potentially be used to predict rapid and sudden shifts toward episodes of uncontrolled substance use.

These studies also foreshadow exciting opportunities for using mobile behavioral tasks in interventions to assess how risk for substance use fluctuates and to specifically deliver treatments at times when they are most needed. In a longitudinal study, Konova et al. [74] used a computer-based risky decision-making task and computational models to track risk tolerance and ambiguity tolerance in patients with opioid use disorder. They found that increased ambiguity tolerance, as extracted by computational modeling, predicted increased prospective opioid use and consequent treatment dropout. Importantly, Konova et al. [74] also varied the rate of sampling risky decision-making (between 1-week and 1-month intervals) and only found the association at the highest sampling rate. They suggested that fine-grained sampling rates could improve task-based prediction of substance use, but note that mobile tasks are necessary to achieve such high sampling rates. In line with this argument, mobile tasks, as discussed in this article, could provide an important tool for clinical risk stratification. To amplify potential, many behavioral tasks could, in principle, also be adapted

to deliver treatments as for the approach-avoidance task [4]. Today, only few of such smartphone-based versions of training tasks have been systematically tested. Bahadoor et al. [75] reviewed 22 controlled trials of mobile applications targeting substance use disorder. Only one of these apps used a task-based training. In the latter study, Crane et al. [76] used a smartphone-based approach-avoidance bias training app and found that – when combined with normative feedback – it significantly reduced consequent alcohol consumption. In sum, future task-based smartphone apps could be used both to stratify risk to deliver treatments when they are most needed and to create novel task-based interventions apps.

Despite the technological advantages of using smartphones to build behavioral tasks, designing smartphone apps comes with unique challenges (for an overview, see Piwek et al. [77]): for example, most researchers do not have the training required to program smartphone apps and hiring dedicated programmers can be expensive, especially for smaller research groups). Data protection is essential when patient-data are sent via the internet to remote servers, often hosted by private companies. Open-source frameworks designed to make programing apps easier could help with these challenges. To our knowledge, no openly available framework exists that helps researchers to create apps that can run behavioral tasks in EMA studies. On the one hand, such frameworks could further improve psychometrics by allowing researchers to create "task batteries" and to more easily compare psychometric properties. On the other hand, it could bring EMA-based behavioral tasks to a larger group of clinical researchers in an open-source manner and thereby foster development of more sustainable EMA-based task research through code and data sharing.

Outlook and Conclusion

Here, we reviewed the current use of behavioral tasks in EMA studies in substance use research. Although currently studies in this evolving field are scarce and several methodological issues have to be overcome, existing findings are promising and may encourage researchers to implement tasks in EMA studies. Nonetheless, the field needs to make important steps forward to deliver promises: First of all, more studies are required that systematically compare and improve psychometric properties of smartphone-based tasks. Second, an increased understanding of how behavioral task measures fluctuate over time is necessary to choose adequate sampling schemes

that balance power to detect effects with the potential burden to participants. Finally, knowledge from EMA-based task studies should be translated to interventions, to deliver interventions to substance users when they are most needed by create novel, mobile, and thereby individualized task-based interventions. Openly available programing frameworks that allow researchers to more easily create EMA-based task studies could help with achieving these goals and give EMA-based task research the best possible shot to fulfill its potential in the coming years.

Conflict of Interest Statement

The authors have no conflicts of interest to declare.

Author Contributions

H.G.Z. and L.D. conducted the literature research and drafted the first version of the manuscript. M.R., U.W.E.P., H.T., M.A.R., A.H., R.J.D., and M.N.S. gave critical feedback, revised, and approved the final manuscript.

References

1 GBD 2016 Alcohol collaborators. Alcohol use and burden for 195 countries and territories, 1990–2016: a systematic analysis for the Global Burden of Disease Study 2016. Lancet. 2018 Sep 22;392(10152):1015–35.

2 Heinz A, Kiefer F, Smolka MN, Endrass T, Beste C, Beck A, et al. Addiction research consortium: losing and regaining control over drug intake (ReCoDe): from trajectories to mechanisms and interventions. Addict Biol. 2020 Mar;25(2):e12866.

3 Berridge KC, Robinson TE, Aldridge JW. Dissecting components of reward: "liking", "wanting", and learning. Curr Opin Pharmacol. 2009 Feb 1;9(1):65–73.

4 Wiers RW, Gladwin TE, Hofmann W, Salemink E, Ridderinkhof KR. Cognitive bias modification and cognitive control training in addiction and related psychopathology: Mechanisms, clinical perspectives, and ways forward. Clin Psychol Sci. 2013 Apr;1(2):192–212.

5 Hildebrandt MK, Dieterich R, Endrass T. Neural correlates of inhibitory control in relation to the degree of substance use and substance-related problems: a systematic review and perspective. Neurosci Biobehav Rev. 2021 Jun 5.

6 Hallgren KA, McCrady BS. Interference in the alcohol Stroop task with college student binge drinkers. J Behav Health. 2013;2(2):112–9.

7 Peeters M, Wiers RW, Monshouwer K, van de Schoot R, Janssen T, Vollebergh WA. Automatic processes in at-risk adolescents: the role of alcohol-approach tendencies and response inhibition in drinking behavior. Addiction. 2012 Nov;107(11):1939–46.

8 Roefs A, Huijding J, Smulders FT, MacLeod CM, de Jong PJ, Wiers RW, et al. Implicit measures of association in psychopathology research. Psychol Bull. 2011 Jan;137(1):149–93.

9 Boffo M, Zerhouni O, Gronau QF, van Beek RJ, Nikolaou K, Marsman M, et al. Cognitive bias modification for behavior change in alcohol and smoking addiction: bayesian meta-analysis of individual participant data. Neuropsychol Rev. 2019 Mar 15;29(1):52–78.

10 Ekhtiari H, Victor TA, Paulus MP. Aberrant decision-making and drug addiction: how strong is the evidence? Curr Opin Behav Sci. 2017 Feb 1;13:25–33.

11 Schwarz N. Why researchers should think "real-time": a cognitive rationale. In: Mehl MR, Conner TS, editors. Handbook of research methods for studying daily life. New York: The Guilford Press; 2012. p. 22–42.

12 Solhan MB, Trull TJ, Jahng S, Wood PK. Clinical assessment of affective instability: comparing EMA indices, questionnaire reports, and retrospective recall. Psychol Assess. 2009 Sep;21(3):425–36.

13 Shiffman S. Ecological momentary assessment (EMA) in studies of substance use. Psychol Assess. 2009 Dec;21(4):486–97.

14 Mehl MR, Conner TS, editors. Handbook of research methods for studying daily life. New York: The Guilford Press; 2012.

15 Trull TJ, Ebner-Priemer U. Ambulatory assessment. Annu Rev Clin Psychol. 2013 Mar 28;9:151–76.

16 Wilhelm P, Perrez M, Pawlik K. Conducting research in daily life: A historical review. In: Mehl MR, Conner TS, editors. Handbook of research methods for studying daily life. New York: The Guilford Press; 2012. p. 22–42.

17 Miller G. The smartphone psychology manifesto. Perspect Psychol Sci. 2012 May;7(3):221–37.

18 Brown HR, Zeidman P, Smittenaar P, Adams RA, McNab F, Rutledge RB, et al. Crowdsourcing for cognitive science: the utility of smartphones. PloS One. 2014 Jul 15;9(7):e100662.

19 Rutledge RB, Skandali N, Dayan P, Dolan RJ. A computational and neural model of momentary subjective well-being. Proc Natl Acad Sci U S A. 2014 Aug 19;111(33):12252–7.

20 Shiffman S, Waters AJ. Negative affect and smoking lapses: a prospective analysis. J Consult Clin Psychol. 2004 Apr;72(2):192.

21 Mereish EH, Padovano HT, Wemm S, Miranda R. Appetitive startle modulation in the human laboratory predicts Cannabis craving in the natural environment. Psychopharmacology. 2018 Jul;235(7):1933–43.

22 Miranda R Jr, Wemm SE, Treloar Padovano H, Carpenter RW, Emery NN, Gray JC, et al. Weaker memory performance exacerbates stress-induced cannabis craving in youths' daily lives. Clin Psychol Sci. 2019 Sep;7(5):1094–108.

23 Treloar Padovano H, Miranda R Jr. How adolescents' working memory abilities relate to their alcohol craving in real-life contexts depends on biological sex. Drug Alcohol Depend. 2021 Apr 1;221:108642.

24 Shiffman S, Li X, Dunbar MS, Tindle HA, Scholl SM, Ferguson SG. Does laboratory cue reactivity correlate with real-world craving and smoking responses to cues? Drug Alcohol Depend. 2015 Oct 1;155:163–9.

25 Begh R, Smith M, Ferguson SG, Shiffman S, Munafò MR, Aveyard P. Association between smoking-related attentional bias and craving measured in the clinic and in the natural environment. Psychol Addict Behav. 2016 Dec;30(8):868–75.

26 Groefsema M, Engels R, Kuntsche E, Smit K, Luijten M. Cognitive biases for social alcohol-related pictures and alcohol use in specific social settings: an event-level study. Clin Exp Res. 2016 Sep;40(9):2001–10.

27 Snelleman M, Schoenmakers TM, van de Mheen D. Attentional bias and approach/avoidance tendencies do not predict relapse or time to relapse in alcohol dependency. Alcohol Clin Exp Res. 2015 Sep;39(9):1734–9.

28 Hendriks H, van de Rest O, Snippe A, Kieboom J, Hogenelst K. Alcohol consumption, drinking patterns, and cognitive performance in young adults: a cross-sectional and longitudinal analysis. Nutrients. 2020 Jan;12(1):200.

29 Muraven M, Baumeister RF. Self-regulation and depletion of limited resources: does self-control resemble a muscle? Psychol Bull. 2000 Mar;126(2):247–59.

30 Inzlicht M, Schmeichel BJ. What is ego depletion? Toward a mechanistic revision of the resource model of self-control. Perspect Psychol Sci. 2012 Sep;7(5):450–63.

31 Hedge C, Powell G, Sumner P. The reliability paradox: why robust cognitive tasks do not produce reliable individual differences. Behav Res Methods. 2018 Jun;50(3):1166–86.

32 Spearman C. The proof and measurement of association between two things. Int J Epidemiol. 2010 Oct 1;39(5):1137–50.

33 Smith A, de Salas K, Lewis I, Schüz B. Developing smartphone apps for behavioural studies: the alcorisk app case study. J Biomed Inform. 2017 Aug 1;72:108–19.

34 Enkavi AZ, Eisenberg IW, Bissett PG, Mazza GL, MacKinnon DP, Marsch LA, et al. Large-scale analysis of test-retest reliabilities of self-regulation measures. Proc Natl Acad Sci U S A. 2019 Mar 19;116(12):5472–7.

35 Shankman SA, Kaiser AJE, Shenberger ER. The importance of examining psychometrics of biomarkers. Biol Psychiatry Cogn Neurosci Neuroimaging. 2020 Apr 1;5(4):379–80.

36 Bouvard A, Dupuy M, Schweitzer P, Revranche M, Fatseas M, Serre F, et al. Feasibility and validity of mobile cognitive testing in patients with substance use disorders and healthy controls. Am J Addict. 2018 Oct; 27(7):553–6.

37 Sliwinski MJ, Mogle JA, Hyun J, Munoz E, Smyth JM, Lipton RB. Reliability and validity of ambulatory cognitive assessments. Assessment. 2018 Jan;25(1):14–30.

38 Pal R, Mendelson J, Clavier O, Baggott MJ, Coyle J, Galloway GP. Development and testing of a smartphone-based cognitive/neuropsychological evaluation system for substance abusers. J Psychoactive Drugs. 2016 Aug 7; 48(4):288–94.

39 Jones A, Tiplady B, Houben K, Nederkoorn C, Field M. Do daily fluctuations in inhibitory control predict alcohol consumption? An ecological momentary assessment study. Psychopharmacology. 2018 May;235(5):1487–96.

40 Spanakis P, Jones A, Field M, Christiansen P. A stroop in the hand is worth two on the laptop: superior reliability of a smartphone based alcohol stroop in the real world. Subst Use Misuse. 2019 Mar 21;54(4):692–8.

41 Nunnally JC, Bernstein IH. Psychometric theory 3E. New York: McGraw-Hill Education; 1994.

42 Lance CE, Butts MM, Michels LC. The sources of four commonly reported cutoff criteria: what did they really say? Organ Res Methods. 2006 Apr;9(2):202–20.

43 Emery NN, Simons JS. The role of affect, emotion management, and attentional bias in young adult drinking: an experience sampling study. Psychopharmacology. 2020 May; 237(5):1557–75.

44 Shahar N, Hauser TU, Moutoussis M, Moran R, Keramati M; Neuroscience In Psychiatry Network Consortium, et al. Improving the reliability of model-based decision-making estimates in the two-stage decision task with reaction-times and drift-diffusion modeling. PLoS Comput Biol. 2019 Feb 13;15(2): e1006803.

45 Haines N, Kvam PD, Irving LH, Smith C, Beauchaine TP, Pitt MA, et al. Theoretically Informed Generative Models Can Advance the Psychological and Brain Sciences: Lessons from the Reliability Paradox 2020; 2020.

46 Waltmann M, Schlagenhauf F, Deserno L. Sufficient reliability of the behavioral and computational read-outs of a probabilistic reversal learning task; 2021.

47 Marhe R, Waters AJ, van de Wetering BJ, Franken IH. Implicit and explicit drug-related cognitions during detoxification treatment are associated with drug relapse: an ecological momentary assessment study. J Consult Clin Psychol. 2013 Feb;81(1):1–12.

48 Suffoletto B, Field M, Chung T. Attentional and approach biases to alcohol cues among young adult drinkers: An ecological momentary assessment study. Exp Clin Psychopharmacol. 2020 Dec;28(6):649–58.

49 MacLean RR, Sofuoglu M, Waters AJ. Naturalistic measurement of dual cue attentional bias in moderate to heavy-drinking smokers: a preliminary investigation. Drug Alcohol Depend. 2020 Apr 1;209:107892.

50 Schmukle SC. Unreliability of the dot probe task. Eur J Pers. 2005 Dec;19(7):595–605.

51 Brown VM, Chen J, Gillan CM, Price RB. Improving the reliability of computational analyses: Model-based planning and its relationship with compulsivity. Biol Psychiatry Cogn Neurosci Neuroimaging. 2020 Jun 1;5(6): 601–9.

52 Sutton RS, Barto AG. Introduction to reinforcement learning. Cambridge: MIT press; 1998.

53 Zech H, Lee Y, Ebner-Priemer U, Garbusow M, Heinz A, Marxen M, et al. Preregistration: do changes in cognitive control and decision making predict changes in drinking behavior? 2020.

54 Krönke KM, Mohr H, Wolff M, Kräplin A, Smolka MN, Bühringer, et al. Real-life self-control is predicted by parietal activity during preference decision making: a brain decoding analysis. Cogn Affect Behav Neurosci. 2021 Jun 1.

55 Krönke KM, Wolff M, Shi Y, Kräplin A, Smolka MN, Bühringer G, et al. Functional connectivity in a triple-network saliency model is associated with real-life self-control. Neuropsychologia. 2020 Dec;149:107667.

56 Wolff M, Enge S, Kräplin A, Krönke KM, Bühringer G, Smolka MN, et al. Chronic stress, executive functioning, and real-life self-control: an experience sampling study. J Pers. 2021 May;89(3):402–21.

57 Krönke KM, Wolff M, Mohr H, Kräplin A, Smolka MN, Bühringer G, et al. Predicting real-life self-control from brain activity encoding the value of anticipated future outcomes. Psychol Sci. 2020 Mar;31(3):268–79.

58 Krönke KM, Wolff M, Mohr H, Kräplin A, Smolka MN, Bühringer G, et al. Monitor yourself! Deficient error-related brain activity predicts real-life self-control failures. Cogn Affect Behav Neurosci. 2018 Aug;18(4):622–37.

59 Wolff M, Krönke KM, Venz J, Kräplin A, Bühringer G, Smolka MN, et al. Action versus state orientation moderates the impact of executive functioning on real-life self-control. J Exp Psychol Gen. 2016 Dec;145(12):1635–53.

60 McKinney TL, Euler MJ, Butner JE. It's about time: the role of temporal variability in improving assessment of executive functioning. Clin Neuropsychol. 2020 May 18;34(4):619–42.

61 Kennedy DO, Veasey RC, Watson AW, Dodd FL, Jones EK, Tiplady B, et al. Vitamins and psychological functioning: a mobile phone assessment of the effects of a B vitamin complex, vitamin C and minerals on cognitive performance and subjective mood and energy. Hum Psychopharmacol. 2011 Jun;26(4–5):338–47.

62 Zech HG, Rotteveel M, van Dijk WW, van Dillen LF. A mobile approach-avoidance task. Behav Res Methods. 2020;52(5):2085–97.

63 Niv Y, Daw ND, Joel D, Dayan P. Tonic dopamine: opportunity costs and the control of response vigor. Psychopharmacology. 2007 Apr;191(3):507–20.

64 Jung Y, Kim J, Awofeso O, Kim H, Regnier F, Bae E. Smartphone-based colorimetric analysis for detection of saliva alcohol concentration. Appl Opt. 2015 Nov 1;54(31):9183–9.

65 Suffoletto B, Gharani P, Chung T, Karimi H. Using phone sensors and an artificial neural network to detect gait changes during drinking episodes in the natural environment. Gait Posture. 2018 Feb 1;60:116–21.

66 Huys QJ, Maia TV, Frank MJ. Computational psychiatry as a bridge from neuroscience to clinical applications. Nat Neurosci. 2016 Mar; 19(3):404–13.

67 Eldar E, Roth C, Dayan P, Dolan RJ. Decodability of reward learning signals predicts mood fluctuations. Curr Biol. 2018 May 7; 28(9):1433–e7.

68 Volkow ND, Wang GJ, Fowler JS, Tomasi D, Telang F, Baler R. Addiction: decreased reward sensitivity and increased expectation sensitivity conspire to overwhelm the brain's control circuit. Bioessays. 2010 Sep;32(9): 748–55.

69 Birch CD, Stewart SH, Wall AM, McKee SA, Eisnor SJ, Theakston JA. Mood-induced increases in alcohol expectancy strength in internally motivated drinkers. Psychol Addict Behav. 2004 Sep;18(3):231.

70 Grant VV, Stewart SH, Birch CD. Impact of positive and anxious mood on implicit alcohol-related cognitions in internally motivated undergraduate drinkers. Addict Behav. 2007 Oct 1;32(10):2226–37.

71 Greenfield SF, Weiss RD, Muenz LR, Vagge LM, Kelly JF, Bello LR, et al. The effect of depression on return to drinking: a prospective study. Arch Gen Psychiatry. 1998 Mar 1; 55(3):259–65.

72 Hogarth L. Addiction is driven by excessive goal-directed drug choice under negative affect: translational critique of habit and compulsion theory. Neuropsychopharmacology. 2020 Apr;45(5):720–35.

73 Witkiewitz K, Villarroel NA. Dynamic association between negative affect and alcohol lapses following alcohol treatment. J Consult Clin Psychol. 2009 Aug;77(4):633.

74 Konova AB, Lopez-Guzman S, Urmanche A, Ross S, Louie K, Rotrosen J, et al. Computational markers of risky decision-making for identification of temporal windows of vulnerability to opioid use in a real-world clinical setting. JAMA Psychiatry. 2020 Apr 1;77(4): 368–77.

75 Bahadoor R, Alexandre JM, Fournet L, Gellé T, Serre F, Auriacombe M. Inventory and analysis of controlled trials of mobile phone applications targeting substance use disorders: a systematic review. Front Psychiatry. 2021 Feb 22;12:142.

76 Crane D, Ubhi HK, Brown J, West R. Relative effectiveness of a full versus reduced version of the "smoke free" mobile application for smoking cessation: an exploratory randomised controlled trial. F1000 Res. 2018:7.

77 Piwek L, Ellis DA, Sally A. Can programming frameworks bring smartphones into the mainstream of psychological science? Front Psychol. 2016 Aug 23;7:1252.