Article:

Spanakis, P., Jones, A., Field, M.J. orcid.org/0000-0002-7790-5559 et al. (1 more author) (2018) A stroop in the hand is worth two on the laptop: superior reliability of a smartphone based alcohol stroop in the real world. Substance Use & Misuse, 54 (4). pp. 692-698. ISSN 1082-6084

https://doi.org/10.1080/10826084.2018.1536716

SUBSTANCE USE AND MISUSE

FILE NOT FOR REVIEW

A Stroop in the hand is worth two on the laptop: Superior reliability of a Smartphone based alcohol Stroop

Running title: Smartphone alcohol Stroop reliability.

Panagiotis Spanakis[a], Andrew Jones[a,b], Matt Field[a,b] and Paul Christiansen[a,b]✉.

[a]*Department of Psychological Sciences, University of Liverpool, Liverpool, UK.*

[b] *UK Centre for Tobacco and Alcohol Studies (UKCTAS),*

✉ Panagiotis Spanakis

School of Psychology

The University of Liverpool

Eleanor Rathbone Building

Bedford Street South

Liverpool L69 7ZA

panagiotis.spanakis@liverpool.ac.uk

tel.: +44 (0) 151 7957510

Running title: Smartphone alcohol Stroop reliability.

This manuscript has not been published elsewhere and it has not been submitted for publication elsewhere.

ABSTRACT

Background. Attentional bias (AB) is the tendency for substance-related stimuli to grab attention. The addiction Stroop task is widely used for measuring AB in the lab, but it has poor reliability and inconsistent predictive validity. Therefore, there is a need to improve the psychometric properties of the task.

Objectives. This study contrasts the internal reliability and predictive validity of a basic (general alcohol words) and an upgraded (personalized alcohol pictures) Stroop task delivered in a neutral room on a laptop computer versus a smartphone application on participants' homes.

Methods. 120 participants recruited in 2016, completed a basic and an upgraded Stroop task. Half the participants completed the tasks on a laptop computer in a neutral university room and the remainder completed the tasks on a smartphone app in their homes. Participants also self-reported their typical alcohol consumption.

Results. Acceptable internal reliability was found for both basic (α = .70) and upgraded (α = .74) Stroop tasks in the home-smartphone condition, but neither had acceptable internal reliability when administered in the neutral room-computer condition (basic α = .49; upgraded α = .58). Participants showed AB toward alcohol words, but not for beer pictures, regardless of condition. None of the indices of AB was associated with individual differences in alcohol involvement, regardless of condition.

Conclusion. The internal reliability of the alcohol Stroop was acceptable when administered on smartphones, in naturalistic settings, rather than on laptop computers, in neutral university rooms.

Keywords: Alcohol; Attentional bias; Reliability; Smart-phone; Stroop; Validity

Introduction

Incentive-motivational models of addiction (e.g. Franken, 2003) suggest that a central factor in the aetiology of substance abuse is the ability of substance-related environmental cues to grab and hold attention (known as attentional bias; AB). Several paradigms have been developed to measure AB for substance-related stimuli, with the addiction Stroop task being one of the most commonly used (for a review see Cox, Fadardi, & Pothos, 2006). In the addiction Stroop, participants are presented with substance-related and control stimuli in different colours and are required to name the colour the stimulus is presented in. It is argued that slower colour naming of substance-related, compared to control stimuli, is indicative of AB, as the substance-related stimuli are grabbing and/or holding attention.

A large number of studies have revealed an association between AB, measured with the addiction Stroop, and substance use/dependence (for a comprehensive review see Field & Cox, 2008). There are, however, numerous examples of studies that fail to show this association (Ryan, 2002; Sharma, Albery, & Cook, 2001; Snelleman, Schoenmakers, & Mheen, 2015) and the clinical utility of these measures is limited (see Christiansen, Schoenmakers, & Field, 2015). One explanation for these inconsistencies might be the poor psychometric properties of the Stroop. For example, Ataya, Adams, Mulings, Cooper, Attwood and Munafo (2012) analysed the internal reliability of addiction Stroop tasks used in nine studies and found that only two had acceptable internal reliability.

Several proposals have been made for the improvement of the psychometric properties of the addiction Stroop task. Personalized stimuli (see Field & Christiansen, 2012) refers to the use of stimuli that reflect participants' drinking preferences. As AB is argued to develop through repeated pairings of environmental cues with substance use, it would be unlikely for an individual to show AB for a type of alcohol they rarely consume or find aversive (e.g. people who only consume beer should not show AB towards whiskey-related stimuli). By tailoring Stroop stimuli to an individuals' conditioning history, reliability should be increased. Indeed, when measuring AB using a visual probe task, Christiansen et al. (2015) found that personalised stimuli produced acceptable levels of internal reliability compared to generalised alcohol stimuli (although no improvements in predictive validity were found). However, a personalised (but not general) card-based alcohol Stroop was found to predict drinking behaviour in undergraduate drinkers (Christiansen & Bloor, 2014). It has also been argued that the psychometric properties of the Stroop could be improved by using pictorial stimuli as they are a closer representation of real cues, in the real world, thereby serving to increase stimulus salience leading to more consistent responding (see relevant discussion in Bruce & Jones, 2004).

Although there have been attempts to explore and improve the psychometric properties of lab-based computerised Stroop tasks, it is notable that recent years have seen increased uptake of ecological momentary assessment (EMA) methods to assess attentional bias in the real world using smart phones and similar devices. Field et al. (2016) argued that repeatedly measuring AB in the real world is critical because AB is a marker of the current motivational state that is itself determined by a range of factors (for example availability, mood, and craving) rather than being a stable trait. Indeed, EMA methods have already been used to deliver substance-related Stroop tasks (e.g. Marhe, Waters, van de Wetering, & Franken, 2013; Waters & Li, 2008; Waters et al., 2014), and smartphone-based attentional bias modification applications are already commercially available. However, despite arguments that smartphone-delivered cognitive tasks should enhance user-device interaction leading to more accurate and meaningful measurements (for examples of cognitive tasks that have been reliably administered on smartphones see Brown et al., 2014; Jones, Tiplady, Houben, Nederkoorn, & Field, 2018) the psychometric properties of smart-phone delivered addiction Stroop tasks outside the lab have yet to be investigated.

The aim of the current study was to explore the internal reliability and predictive validity of a basic Stroop task (i.e. general alcohol words) and an upgraded Stroop that would combine the improvement suggestions mentioned above (i.e. personalized alcohol stimuli presented as pictures). We also compared, on a between subjects basis, the reliability of these tasks when completed in a neutral university room on a computer (standard methodology) or in participants' own homes on a smartphone, (standard EMA methodology, see for example Marhe et al., 2013; Waters, Marhe, & Franken, 2012). We hypothesised that the upgraded Stroop would show acceptable reliability in both conditions

We also predicted that participants should demonstrate a robust AB for alcohol in both versions of the Stroop and that individual differences in alcohol use would be positively correlated with AB.

Methods

Participants

One hundred and twenty participants (61 female) were recruited from the University of Liverpool and local community via advertisements and word of mouth. Inclusion criteria were; fluent English speaking, regular social drinker (consuming at least one alcoholic drink in a typical week), and beer being their preferred alcoholic drink. Participants were excluded if they had any current or previous alcohol abuse or dependence diagnosis, were pregnant or breastfeeding, or colour-blind. Participant characteristics are shown in Table 1. All participants provided informed consent prior to participation and the study was approved by the University of Liverpool Ethics committee.

Materials

Questionnaires

*Time Line Follow Back (TLFB; Sobell & Sobell, 1990).*The TLFB self-report questionnaire was used to retrospectively assess alcohol consumption. Participants had to estimate the number of alcohol units consumed over the preceding 7 days (one UK unit = 8g of alcohol). TLFB self-report questionnaires has previously shown high test-retest reliability (r = .98) for total number of reported drinks in a sample of college students (Sobell, Sobell, Klajner, Pavan, & Basian, 1986)

The Alcohol Use Disorders Identification Test (AUDIT; Saunders, Aasland, Babor, De la Fuente, & Grant, 1993). The AUDIT self-report questionnaire was used to assess hazardous drinking. It consists of ten fixed-response questions regarding alcohol consumption and consequences of drinking. Scores range between 0 and 40, with scores >8 indicating hazardous alcohol use. Self-reported AUDIT questionnaire has previously shown high internal reliability (α = .82) in a sample of college students (Shields, Guttmannova, & Caruso, 2004).

Alcohol Stroop Tasks

Stimuli.

General words. 11 general alcohol-related words (e.g. pub, beer, wine, etc.) and 11 control words (e.g. bog, ravine, valley, etc.) were presented in three different font colours (blue, green, red). Words were matched in terms of word length and word frequency in the English language (see Sharma et al., 2001).

Personalized pictures. 11 beer-related pictures (bottles or cans of beer) and 11 control pictures (bottles or cans of soft drinks) were recoloured in three different colours (blue, green, red). As preference for beer was an inclusion criterion, beer pictures represent a personalised stimulus. We chose to use pictures of cans and bottles of known beer brands as they are easily and widely recognizable as an alcoholic drink, even after recolouring. Pictures selected for this study were matched in terms of presentation features (e.g. size and packaging).

Task description.

Each trial commenced with a fixation dot, presented centrally for 500ms. Following this, a stimulus (alcohol-related or control) was presented on the centre of the screen in either red, blue, or green font (word Stroop) or with a red, blue or green filter (pictorial Stroop). Below the stimuli, three buttons with the three different colour names appeared on the screen, with their position changing on each trial. Participants were instructed to indicate the colour of the stimulus by mouse clicking (computer condition) or screen tapping (smart-phone condition) the corresponding button, as fast and as accurately as possible. Stimuli remained on the screen until a response was given or after a timeout period of 3000ms. If the response was incorrect or slower than 3000ms, a red “X” appeared on the screen for 500ms. The inter-trial interval was 500ms. Each stimulus appeared once in each colour and stimuli were presented in two blocks (alcohol and control, 33 trials per block), with a five-second inter-block pause. Blocked presentation was chosen to avoid any interference carry over effects (see Waters, Sayette, & Wertz, 2003). The task was programmed with Open Sesame software (version 2.9 series; Mathôt, Schreij, & Theeuwes, 2012).

Procedure

All participants provided informed consent before providing demographic information and self-completing the TLFB and AUDIT questionnaires, with a researcher being present to provide instructions and assist them if needed. They then completed two versions of the alcohol Stroop task, a basic with alcohol words generally related to alcohol and an upgraded with pictures of participants’ preferred drink (beer pictures), in a counterbalanced order. In both tasks, blocks were presented in a fixed order, with alcohol blocks presented first. Participants were recruited in two blocks; the first block (n=60) completed the Stroop tasks on a standard laptop computer in a neutral control room in the university campus (i.e. empty class rooms, library rooms, etc) and the second block (n = 60) were loaned a smartphone (Samsung Galaxy Ace GT-S5830i) to complete the task in their homes, after the task was thoroughly explained to them. These conditions were created to resemble the common methodologies used either in laboratory experiments, where participants usually complete the Stroop task on a standard desktop/laptop computer in a neutral laboratory room, and in EMA studies where they usually complete the task in a portable, handheld device in a naturalistic environment. At the end of the study participants were debriefed.

Data reduction and analysis

Before calculating reaction times, responses that were incorrect, faster than 200ms, slower than 2000ms or three standard deviations above the individual mean response time were removed (see Schoenmakers, Wiers, & Field, 2008). Following that, mean reaction time on control trials was subtracted from mean reaction time on alcohol trials; positive scores are indicative of greater AB towards alcohol-related cues.

Internal reliability of the tasks was calculated with Cronbach's α, where α ≥ .70 was considered acceptable (Kline, 1999). For each pair of stimuli (alcohol stimulus and its matched control) a separate AB score was calculated. First, we calculated the average response time (rt) across all colour presentation for each alcohol stimulus and its paired control and then we subtracted the control rt from the alcohol rt. Given that each Stroop task included 11 pairs of alcohol and control stimuli, Cronbach's α reflected the internal consistency among those 11 stimuli-specific AB scores. Differences in AB magnitude were examined with a 2x2x2 mixed ANOVA with Stroop type (basic, upgraded) and stimulus type (alcohol-related, control) as within subjects' factors and condition (neutral room-computer, home-smartphone) as the between subjects' factor. For predictive validity, alcohol use involvement was calculated as the sum of the standardised TLFB and AUDIT scores (see Christiansen & Bloor, 2014). Due to multicollinearity problems in the interaction terms between AB scores and condition (VIFs > 10), a separate regression was run for each condition. The regression models examined the association between AB scores and alcohol involvement, after controlling for age and gender. P values were corrected for multiple test by multiplying them by two.

[Table 1 here]

Results

Internal reliability

An acceptable level of reliability was achieved for home-smartphone based Stroop tasks both in the basic (alcohol words, α=.70) and upgraded type (beer pictures, α=.74). The neutral room-computer based Stroop task did not have acceptable reliability, neither in the basic (α=.49) or upgraded type (α=.58).

Differences in magnitude of alcohol AB

A mixed design ANOVA revealed a main effect of Stroop type (basic or upgraded) (F (1, 118) = 4.61, p =.034, η_p^2 = .04) with participants responding slower to pictures compared to words. There was also a main effect of stimulus type (alcohol or control) (F (1, 118) = 23.22, p <.001, η_p^2 = .16) with slower responses to alcohol stimuli compared to control.

There was also a main effect of condition (neutral room-computer or home-smartphone) (*F* (1, 118) = 15.34, *p* <.001, η_p^2 = .12) with slower responses on the neutral room-computer compared to the home-smartphone.

A significant three-way interaction was found between stimulus type, Stroop type and condition (*F* (1, 118) = 5.43, *p* = .021, η_p^2 = .04). Within the neutral room-computer condition, participants responded slower to alcohol words compared to control in the basic Stroop (*t*(59) = 7.97, *p* <.001, *d*=.52) but there was no difference between beer and control pictures in the upgraded Stroop (*t*(59) = -0.39, *p* = .70, *d* = - .03). The same pattern was found within the home-smartphone condition, with participants responding slower to alcohol words compared to control in the basic Stroop (*t(*59) = 2.93, *p* =.005, *d*=.22) but showing no difference between beer and control pictures in the upgraded Stroop (*t*(59) = -0.40, *p* = .69, *d* = -.02).

Predictive validity: Regression between alcohol AB and alcohol involvement.

Regardless of condition (neutral room-computer or home-smartphone), neither AB for words (basic type) nor AB for beer pictures (upgraded type) predicted individual differences in alcohol involvement (*ps*>.05), see table 2.

[Table 2 here]

Post-hoc analysis: Comparison of RTs across stimulus category.

Given that attentional bias for alcohol related stimuli was found only for alcohol words but not beer pictures, we examined the role of the control stimuli used in each case. Reaction times were compared between alcohol-words and beer pictures and between control words (environmental features) and control pictures (soft drinks). Participants were found to be significantly slower when reacting to control pictures compared to control words ($t(119) = -5.36$, $p < .001$, $d = -.30$), although no difference was found between alcohol words and beer pictures ($t(119) = 1.28$, $p = .202$, $d = .08$).

Discussion

In the current study we contrasted the psychometric properties and predictive validity of a basic (general alcohol words) and an upgraded (personalized pictures) Stroop task, administered on a standard computer in a neutral room and on a smartphone in participants' homes. Findings partially supported our hypotheses. We found that the alcohol Stroop had acceptable reliability only when administered on the smartphone in naturalistic environments and not on the computer in a neutral university room, regardless of whether participants were exposed to words generally related to alcohol (basic type) or personalized pictures of beer (upgraded type). Also, regardless of whether they were assessed on the home-smartphone or the neutral room-computer condition, participants showed attentional bias for general alcohol words but not for the personalized beer pictures. Finally, there was no association between any index of AB and individual differences in alcohol consumption, suggesting poor predictive validity regardless of stimulus type and delivery format.

Acceptable levels of reliability were observed only when the Stroop was administered in people's homes on smartphones, showing that this method could be a better alternative to the conventional computer measurements of AB in the laboratory. Although this supports the methodology used in many EMA studies, where Stroop task is administered on a smartphone/hand-held computer in a naturalistic environment, we cannot unequivocally determine whether this improved reliability is attributed to the ubiquitous nature of smartphones or the ecologically valid environment of homes or a combination of both. It is argued that constant use of smartphones enhances the user-device interaction, ultimately improving measurements taken through them (for examples of cognitive tasks that have been reliably administered on smartphones see Brown et al., 2014; Jones et al., 2018). However, factors related with the underlying motivational state of AB (e.g. alcohol related cues, substance availability, for a review see Field et al., 2016) are more prominent in participants homes, which may also influence reliability of measurement.

Moreover, regardless of environment, Stroop interference for alcohol-related stimuli was found only for the basic and not the upgraded Stroop. The most likely explanation for this failure is choice of the control stimuli. Specifically, beer and soft drink pictures are both appetitive stimuli, in contrast to alcohol and environment related words. As shown in the post-hoc analysis, the appetitive nature of control pictures in the upgraded Stroop may have produced an attentional bias towards them, whereas there was less attention towards non-appetitive environmental words in the basic Stroop. This could reduce the magnitude of the difference in reaction times between the two types of stimuli. It is notable that it is necessary for a beer pictorial Stroop to use soft drink pictures as control stimuli, rather than environmental pictures, as environmental pictures cannot be matched to the beer pictures on perceptual qualities. Neutral, non-appetitive words were used in the word Stroop task to be consistent with previous research (particularly those using EMA; see Marhe et al., 2013), and also due to practical difficulties matching soft drinks words (avoiding brand names) to alcohol words. Another factor that could have affected the efficacy of the upgraded Stroop is strong pre-existing associations between colour and branding (e.g. real Sprite bottles are always green, see Cox et al., 2006). Also, colour filters applied on the pictures may mean that the stimuli are less like those previously experienced by participants in the real world, thus reducing their incentive salience. Future research could explore using (a) appetitive control words and (b) coloured outlines around pictures (see for example cocaine-related Stroop tasks by Hester, Dixon, & Garavan, 2006), in order to explore optimum formats of Stroop tasks across different methods of delivery.

Despite the increase in internal reliability, no measure of AB was associated with alcohol involvement, adding to the equivocal literature (see Christiansen, Schoenmakers, et al., 2015). One possible explanation is that AB is actually a state characteristic which is more likely to be associated with alcohol use in the short term but less likely to be associated with alcohol behaviours in the long term, let alone retrospectively (see Christiansen, Schoenmakers, et al., 2015; Field et al., 2016; Marhe et al., 2013). Future studies could utilize (reliable) smartphone based naturalistic assessments of AB to examine the association between AB and alcohol involvement in the short-term.

Our study had a number of limitations. First, for practical reasons, participants were allocated to conditions in blocks of 60, rather than randomly. However, the same recruitment procedures were followed in both blocks and participants were tested during the same period of the year. Second, regarding personalized stimuli, the current study only recruited beer drinkers due to limitations in developing multiple app versions. Although this methodology still offers meaningful comparisons, we would encourage future studies to tailor personalized stimuli to each participant.

Third, pictorial and personalized stimuli were confounded in the form of beer pictures, rather than examined separately. Although we did not find any reliability or validity differences between the basic and upgraded Stroop, future studies could disentangle pictorial and personalized features to examine the effects of each on the psychometric properties of the alcohol Stroop task. Fourth, as mentioned above, smartphone administration and naturalistic environment were also confounded in a single condition. Although this combination seems pragmatically appropriate, due to the portable and ubiquitous nature of mobile devices, future research should examine the reliability of smartphone-based Stroop in a laboratory and naturalistic environment separately, in order to disentangle the environmental effect from the mean of administration effect. We would also encourage researchers to use geo-location (e.g. the phones GPS) and ask participants for details of the environment (e.g. private/public, presence of alcohol cues, etc) to identify environmental factors that could affect the test's psychometric properties.

Finally, reaction times in the Stroop task appear larger compared to other published alcohol-Stroop task (e.g. Cox, Yeates, & Regan, 1999; Sharma et al., 2001). This difference might be due to the way in which responses were recorded (mouse-clicking or screen-tapping vs pressing buttons; see Lin & Wu, 2013) and due to different number of trials (60 vs 100 and over – participants in longer task might become faster as they get more practice; see Cox et al., 2006). Despite this, the magnitude of reaction time differences between alcohol and control stimuli should not have been affected, meaning that our inferences based on AB scores would remain valid.

Overall, our findings suggest that smartphone-based Stroop task administered in naturalistic environments can have acceptable reliability, although caution should be exerted regarding the choice of control stimuli and the predictive validity of the measurements. We would encourage researchers to analyse reliability of tasks as a matter of course, so that a clear evidence base for the optimal formats of Stroop task, in terms of reliability and predictive validity, can be developed.

Declaration of interest: The authors report no conflicts of interest. The authors alone are responsible for the content and writing of the paper.

References

Ataya, A. F., Adams, S., Mullings, E., Cooper, R. M., Attwood, A. S., & Munafò, M. R. (2012). Internal reliability of measures of substance-related cognitive bias. *Drug and alcohol dependence, 121*(1), 148-151.

Brown, H. R., Zeidman, P., Smittenaar, P., Adams, R. A., McNab, F., Rutledge, R. B., & Dolan, R. J. (2014). Crowdsourcing for Cognitive Science – The Utility of Smartphones. *PLOS ONE, 9*(7), 1-9.

Bruce, G., & Jones, B. T. (2004). A pictorial Stroop paradigm reveals an alcohol attentional bias in heavier compared to lighter social drinkers. *Journal of Psychopharmacology, 18*(4), 527-533.

Christiansen, P., & Bloor, J. F. (2014). Individualised but not general alcohol Stroop predicts alcohol use. *Drug and alcohol dependence, 134*, 410-413.

Christiansen, P., Mansfield, R., Duckworth, J., Field, M., & Jones, A. (2015). Internal reliability of the alcohol-related visual probe task is increased by utilising personalised stimuli and eye-tracking. *Drug and alcohol dependence*. *155*, 170-174.

Christiansen, P., Schoenmakers, T. M., & Field, M. (2015). Less than meets the eye: reappraising the clinical relevance of attentional bias in addiction. *Addictive behaviors, 44*, 43-50.

Cox, W. M., Fadardi, J. S., & Pothos, E. M. (2006). The addiction-stroop test: Theoretical considerations and procedural recommendations. *Psychological bulletin, 132*(3), 443-476.

Cox, W. M., Yeates, G. N., & Regan, C. M. (1999). Effects of alcohol cues on cognitive processing in heavy and light drinkers. *Drug and alcohol dependence, 55*(1), 85-89.

Field, M., & Christiansen, P. (2012). Commentary on,'Internal reliability of measures of substance-related cognitive bias'. *Drug and alcohol dependence, 124*(3), 189-190.

Field, M., & Cox, W. M. (2008). Attentional bias in addictive behaviors: a review of its development, causes, and consequences. *Drug and alcohol dependence, 97*(1), 1-20.

Field, M., Werthmann, J., Franken, I., Hofmann, W., Hogarth, L., & Roefs, A. (2016). The role of attentional bias in obesity and addiction. *Health Psychology, 35*(8), 767-780.

Franken, I. H. (2003). Drug craving and addiction: integrating psychological and neuropsychopharmacological approaches. *Progress in Neuro-Psychopharmacology and Biological Psychiatry, 27*(4), 563-579.

Hester, R., Dixon, V., & Garavan, H. (2006). A consistent attentional bias for drug-related material in active cocaine users across word and picture versions of the emotional Stroop task. *Drug and alcohol dependence, 81*(3), 251-257.

Jones, A., Tiplady, B., Houben, K., Nederkoorn, C., & Field, M. (2018). Do daily fluctuations in inhibitory control predict alcohol consumption? An ecological momentary assessment study. *Psychopharmacology, 235*(5), 1487-1496.

Kline, P. (1999). The Handbook of Psychological Testing 2 edition Routledge. *London, UK*.

Lin, C.-J., & Wu, C. (2013). Reactions, accuracy and response complexity of numerical typing on touch screens. *Ergonomics, 56*(5), 818-831.

Marhe, R., Waters, A. J., van de Wetering, B. J. M., & Franken, I. H. A. (2013). Implicit and explicit drug-related cognitions during detoxification treatment are associated with drug relapse: an ecological momentary assessment study. *Journal of consulting and clinical psychology, 81*(1), 1-12.

Mathôt, S., Schreij, D., & Theeuwes, J. (2012). OpenSesame: An open-source, graphical experiment builder for the social sciences. *Behavior research methods, 44*(2), 314-324.

Ryan, F. (2002). Attentional bias and alcohol dependence: A controlled study using the modified Stroop paradigm. *Addictive behaviors, 27*(4), 471-482.

Saunders, J. B., Aasland, O. G., Babor, T. F., De la Fuente, J. R., & Grant, M. (1993). Development of the alcohol use disorders identification test (AUDIT). WHO collaborative project on early detection of persons with harmful alcohol consumption-II. *Addiction, 88*, 791-791.

Schoenmakers, T., Wiers, R. W., & Field, M. (2008). Effects of a low dose of alcohol on cognitive biases and craving in heavy drinkers. *Psychopharmacology, 197*(1), 169-178.

Sharma, D., Albery, I. P., & Cook, C. (2001). Selective attentional bias to alcohol related stimuli in problem drinkers and non-problem drinkers. *Addiction, 96*(2), 285-295.

Shields, A. L., Guttmannova, K., & Caruso, J. C. (2004). An Examination of the Factor Structure of the Alcohol Use Disorders Identification Test in Two High-Risk Samples. *Substance Use & Misuse, 39*(7), 1161-1182.

Snelleman, M., Schoenmakers, T. M., & Mheen, D. (2015). Attentional bias and approach/avoidance tendencies do not predict relapse or time to relapse in alcohol dependency. *Alcoholism: Clinical and Experimental Research, 39*(9), 1734-1739.

Sobell, L. C., & Sobell, M. B. (1990). Self-report issues in alcohol abuse: State of the art and future directions. *Behavioral Assessment, 12*, 77-90.

Sobell, M. B., Sobell, L. C., Klajner, F., Pavan, D., & Basian, E. (1986). The reliability of a timeline method for assessing normal drinker college students' recent drinking history: utility for alcohol research. *Addictive behaviors, 11*(2), 149-161.

Waters, A., Sayette, M., & Wertz, J. (2003). Carry-over effects can modulate emotional Stroop effects. *Cognition and Emotion, 17*(3), 501-509.

Waters, A. J., & Li, Y. (2008). Evaluating the utility of administering a reaction time task in an ecological momentary assessment study. *Psychopharmacology, 197*(1), 25-35.

Waters, A. J., Marhe, R., & Franken, I. H. (2012). Attentional bias to drug cues is elevated before and during temptations to use heroin and cocaine. *Psychopharmacology, 219*(3), 909-921.

Waters, A. J., Szeto, E. H., Wetter, D. W., Cinciripini, P. M., Robinson, J. D., & Li, Y. (2014). Cognition and craving during smoking cessation: an ecological momentary assessment study. *nicotine & tobacco research, 16*(Suppl 2), S111-S118.

Glossary:

Attentional bias (AB): The involuntary and automatic attraction of attention by cues in the environment related to the substance of use.

Ecological Momentary Assessment (EMA): A methodology to measure behavioural phenomena repeatedly, as they happen and in the natural environment where they occur.

Internal Reliability: A reliability measure based on internal consistency, which describes the extent to which the items of a test measure the same underlying construct and hence correlate highly to each other.

Predictive Validity: The extent to which a test can predict the results of another related test in the future.

Table 1: Sample characteristics (values are means ±SD, apart from Sex for which they represent frequencies)

	Neutral room-computer condition	Home-smartphone condition	Sample
Sex (F:M)	32:28	29:31	61:59
Age	22.28 (±7.28)	23.92 (±9.42)	23.10 (±8.42)
AUDIT[a]	15.27 (±5.49)	15.33(±6.77)	15.30 (±6.14)
Weekly consumption[b] (units)	24.41 (±18.83)	24.82 (±19.69)	24.61 (±19.19)
Reaction time (RT) for alcohol-worlds (sec)	1325.56 (±188.42)	1167.26 (±174.98)	1246.41 (±197.74)
RT for control-words (sec)	1230.48 (±174.13)	1127.83 (±180.44)	1179.16 (±183.94)
RT for beer-pictures (sec)	1282.81 (±164.67)	1178.83 (±179.21)	1230.82 (±179.14)
RT for soft drinks-pictures (sec)	1287.96 (±176.55)	1183.38 (±190.70)	1235.67 (±190.37)

a.AUDIT = Alcohol use disorders identification test scores range between 0 (minimum) and 40 (maximum). b.7-day unit consumption, one UK unit = 8g of alcohol.

Table 2. Regressions showing the association between attentional bias and alcohol involvement after controlling for age and gender.

		β	p[a]	F	p[a]
Neutral room-computer	Step 1				
	Gender	-.07	1	1.04	.72
	Age	-.17	.42		
	Step 2				
	Gender	-.08	1	.91	.92
	Age	-.18	.34		
	AB alcohol-words Stroop (ms)	-.10	.88		
	AB beer-pictures Stroop (ms)	.12	.76		
Home-smartphone	Step 1				
	Gender	-.01	1	.08	1
	Age	-.05	1		
	Step 2				
	Gender	-.03	1	.10	1
	Age	-.01	1		
	AB alcohol-world Stroop (ms)	-.07	1		
	AB beer-pictures Stroop (ms)	.04	1		

a. p values reported here have been doubled to correct for multiple test. Doubled p values have been capped to 1, as in reality p values cannot possibly exceed that number.

(s1) Supplementary Material

P values for the comparison of Cronbach's αs.

Cronbach's αs were compared across conditions and Stroop types using the cocron R package (Diedenhofen & Musch, 2016) to identify meaningful differences. Results are presented in table S1.

Table S1. Reliability comparisons among conditions and Stroop types

Stroop tasks		Cronbach's αs		χ^2 (df)	p
1	2	1	2		
Basic computer	Upgraded computer	.49	.58	0.42 (1)	.52
	Basic smartphone		.70	3.19 (1)	.07
	Upgraded smartphone		.74	5.37 (1)	**.02**
Upgraded computer	Basic smartphone	.58	.70	1.30 (1)	.25
	Upgraded smartphone		.74	2.68	.10
Basic smartphone	Upgraded smartphone	.70	.74	0.25 (1)	.62

A significant increase in reliability is only observed from the basic Stroop in the neutral-room computer condition to the upgraded Stroop in the home-smartphone condition. For all other comparisons, even when the increase is crossing the level of acceptable reliability (α = .70) the difference is not significant.