Taylor & Francis
Taylor & Francis Group

The American Journal of Drug and Alcohol Abuse
Encompassing All Addictive Disorders

ISSN: 0095-2990 (Print) 1097-9891 (Online) Journal homepage: http://www.tandfonline.com/loi/iada20

American Alcohol Photo Stimuli (AAPS): A standardized set of alcohol and matched non-alcohol images

Christopher S. Stauffer MD, Lily Dobberteen BA & Joshua D. Woolley MD, PhD

THE AMERICAN JOURNAL OF DRUG AND ALCOHOL ABUSE
http://dx.doi.org/10.1080/00952990.2016.1253093

American Alcohol Photo Stimuli (AAPS): A standardized set of alcohol and matched non-alcohol images

Christopher S. Stauffer, MD [a,b], Lily Dobberteen, BA[b], and Joshua D. Woolley, MD, PhD[a,b]

[a]Department of Psychiatry, University of California San Francisco, San Francisco, CA, USA; [b]Department of Mental Health, San Francisco Veterans Affairs Medical Center, San Francisco, CA, USA

ABSTRACT

Background: Photographic stimuli are commonly used to assess cue reactivity in the research and treatment of alcohol use disorder. The stimuli used are often non-standardized, not properly validated, and poorly controlled. There are no previously published, validated, American-relevant sets of alcohol images created in a standardized fashion. *Objectives*: We aimed to: 1) make available a standardized, matched set of photographic alcohol and non-alcohol beverage stimuli, 2) establish face validity, the extent to which the stimuli are subjectively viewed as what they are purported to be, and 3) establish construct validity, the degree to which a test measures what it claims to be measuring. *Methods*: We produced a standardized set of 36 images consisting of American alcohol and non-alcohol beverages matched for basic color, form, and complexity. A total of 178 participants (95 male, 82 female, 1 genderqueer) rated each image for appetitiveness. An arrow-probe task, in which matched pairs were categorized after being presented for 200 ms, assessed face validity. Criteria for construct validity were met if variation in AUDIT scores were associated with variation in performance on tasks during alcohol image presentation. *Results*: Overall, images were categorized with >90% accuracy. Participants' AUDIT scores correlated significantly with alcohol "want" and "like" ratings [$r(176) = 0.27$, $p = <0.001$; $r(176) = 0.36$, $p = <0.001$] and arrow-probe latency [$r(176) = -0.22$, $p = 0.004$], but not with non-alcohol outcomes. Furthermore, appetitive ratings and arrow-probe latency for alcohol, but not non-alcohol, differed significantly for heavy versus light drinkers. *Conclusion*: Our image set provides valid and reliable alcohol stimuli for both explicit and implicit tests of cue reactivity. The use of standardized, validated, reliable image sets may improve consistency across research and treatment paradigms.

ARTICLE HISTORY
Received 3 February 2016
Revised 20 October 2016
Accepted 23 October 2016

KEYWORDS
Addiction; alcohol; stimuli; picture; photo; implicit; validation; beverage

Introduction

The presentation of conditioned stimuli (or cues) has been increasingly utilized to assess cue reactivity in the research and treatment of alcohol use disorder (1). The methods reported in the cue reactivity literature are highly heterogeneous (2), with cue presentation modalities ranging from imaginal (e.g., conjuring imagined situations associated with past drinking) to visual (e.g., viewing pictures or videos of alcohol or alcohol-related words) to *in vivo* (e.g., pouring, smelling, or even drinking alcohol of choice). Photographic cues are commonly preferred for their simplicity as well as suitability for the constraints of laboratory paradigms. In previous work on food cue reactivity and craving, photographic cues of food elicit stronger responses than word cues (3) or olfactory cues (4) and have similar effect sizes as exposure to real food (4). Research comparing cue presentation modalities for alcohol are lacking, but addiction research clearly demonstrates that photographic cues of alcohol-related stimuli elicit craving and physiological responses (5,6). Furthermore, alcohol cue reactivity predicts use-related outcomes. For example, implicit approach bias toward alcohol in university students predicts frequency and magnitude of heavy drinking (6), and cue-induced striatal activation seen on functional magnetic resonance imaging in people with alcohol use disorder can predict relapse and subsequent alcohol intake (7). Alcohol cue reactivity is also used to tailor interventions and evaluate the efficacy of treatment programs (2,7,8). For example, naltrexone is suggested to work better at reducing alcohol use in a subgroup of patients with higher cue-induced striatal reactivity (9,10). The photographic stimuli used in these paradigms are often non-standardized, not properly validated, and poorly controlled. In this paper, we validate an original standardized set of photographic cues of alcohol beverages with matched control images of non-

CONTACT Christopher S. Stauffer, MD BandLab@ucsf.edu Department of Mental Health, San Francisco Veterans Affairs Medical Center, 4150 Clement St (116-C), San Francisco, CA 94121, USA.

For complimentary access to AAPS, interested researchers can contact us via e-mail.
Supplemental data for this article can be accessed at publisher's website.

alcohol beverages for use in research and clinical paradigms.

It is important to standardize stimulus sets in order to reduce noise from factors not related to the target cue (i.e., variations in background color or context can confound the assessment of attentional bias for alcohol versus non-alcohol beverages) (11). Standardization involves attention to physical image characteristics (e.g., color, size, contrast, brightness, perceptual complexity), affective characteristics (e.g., appetitiveness, valence, arousal) (12), and contextual characteristics (e.g., bland background versus real-life scenes, inclusion of people versus only non-human objects) (13). Very few rigorously standardized and validated image sets exist for use in alcohol cue reactivity paradigms. Pulido et al. (14) validated a set of alcohol and non-alcohol beverage pictures; however, although the images were matched pairwise on subjective and objective measures of physical image and affective characteristics, they were not created in a standardized fashion. Multiple images were collected from a variety of sources, and matched sets were created *post hoc*. While not an uncommon method in creating alcohol cues, this leads to a set of stimuli containing a diversity of unstandardized contextual factors.

The most expertly standardized photographic stimulus sets are created *de novo* with careful attention to the uniformity of physical and contextual image characteristics. The Amsterdam Beverage Picture Set (ABPS) was created in such a way (13). The ABPS is comprised of unmatched alcohol and non-alcohol beverage photos on white backgrounds. Each beverage is displayed in different contexts, both beverage alone and beverage interacting with a person (being served, opened, or consumed). However, the ABPS was created by Dutch researchers and thus contains beverage brands not readily recognizable in a cross-cultural context. A significant effect of culture was found when comparing North American and German-speaking individuals' ratings of "food-pics", a similarly well-standardized and validated set of food photos created by German researchers (15). No such formal analyses were published for the ABPS, but the authors did acknowledge the need for cultural adaptations of their stimulus set.

To this end, we aimed to: 1) make available a standardized, matched set of photographic alcohol and non-alcohol beverage stimuli with cultural relevance to North America, 2) establish face validity by assessing the degree to which images are subjectively viewed as alcohol versus non-alcohol beverages, and 3) establish construct validity (the degree to which a test measures what it claims to be measuring) for use in clinical and research assessments.

We produced a set of images consisting of beverages chosen for North American cultural relevance, henceforth collectively referred to as the American Alcohol Photo Stimuli (AAPS). AAPS consists of photos of alcohol beverages (n = 18) and non-alcohol beverages (n = 18) matched for basic color, form, and complexity (see Figure 1 for example). A professional product photographer standardized physical image characteristics, such as brightness, contrast, saturation, and sharpness using studio lighting and post-production image manipulation.

For contextual standardization, we opted for beverages on a bland white background. Bottles or cans were opened prior to being photographed to elicit maximum appetitive response, and, where appropriate, a portion of the beverage was poured into a glass and garnished. No other contextual cues are present in the photos. Pronk et al. (13) found no difference in self-reported urge to drink between beverages in isolation and beverages depicted with a person. However, when determining if a photo was an alcohol or non-alcohol beverage, participants responded significantly faster to alcohol images than non-alcohol images when the beverage was in isolation, while there was no difference in reaction time when there was a person in the photo. Thus, AAPS is comprised of photographed beverages in a simple context with a standardized white background, making it most suited to measure implicit cognitive processes that typically use reaction time as the primary outcome.

Stacy and Wiers (16) referred to three broad classes of implicit cognitive processes assumed to underlie the development and maintenance of addictive behaviors: (a) attentional bias for a substance, (b) memory associations related to the substance, (c) action tendencies triggered by the substance (approach or avoidance). Addiction-related implicit cognition tasks include: visual probe tasks, implicit association tasks (IAT) (17), approach-avoidance tasks (AAT), and go/no-go tasks. Compared to measures of self-reported cue reactivity (e.g., craving or impulse to use), implicit cognition is less prone to reporting bias. Moreover, implicit cognition is more efficient than recording and analyzing psychophysiological or imaging data, although these can be used in combination with implicit cognition tasks. Word cues can also be used in implicit cognition tasks, but photographic cues are more accessible to a broader population of research participants, particularly participants with trouble reading or understanding English (17), cognitive impairment, or psychopathology. Although, among heavy drinkers, self-reported craving (18) and physiologic arousal (19) are stronger for beverage photos in a drinking context versus a bland background, tasks measuring alcohol-related implicit cognition are most effective when using images with bland backgrounds (11). Thus, AAPS' lack

Figure 1. Example of matched alcohol (left, Tanqueray gin) and non-alcohol (right, Perrier mineral water) images from the American Alcohol Photo Stimuli (AAPS) set.

of contextual cues and use of a white background supports our goal of creating optimal stimuli for tasks assessing alcohol-related implicit cognitive processes.

Our initial step in validating AAPS was to measure face validity by determining if our alcohol and non-alcohol images are readily identifiable as such. We also wanted to ensure that matched pairs are distinguishable when displayed side-by-side briefly, typically 200-500 milliseconds (ms) in standard tests of implicit cognition. We used an arrow-probe task, based on a validation study of images of tools versus images of matched non-tool objects (20). We hypothesized that matched pairs would be distinguishable with accuracy at or above 89.9%, which was the percent accuracy demonstrated in the Verma et al. (20) task.

In order to establish construct validity for the use of AAPS in clinical and research assessments of implicit cognitive processes, we followed Pronk et al.'s (13) lead by using Borsboom et al.'s (21) definition of validity, which states: a test is valid for measuring an attribute if (a) the attribute exists and (b) variations in the attribute causally produce variations in the measurement outcomes. We aimed to demonstrate that variations in alcohol use will lead to variations in appetitive ratings as well as accuracy and reaction time (latency) in a test of implicit attentional bias. It is well-established that images of alcohol elicit an appetitive response, and appetitive ratings are correlated with excessive drinking (16). Furthermore, reaction time outcomes from tasks of alcohol-related implicit cognition can predict drinking behavior (6). We hypothesized that scores on the Alcohol Use Disorders Identification Test (AUDIT), a marker of alcohol use, would significantly positively correlate with "want" and "like" ratings for alcohol images, but not non-alcohol images. We further hypothesized that AUDIT scores would significantly positively correlate with accuracy and negatively correlate with latency in our arrow-probe task (i.e., higher AUDIT would predict greater accuracy and shorter reaction time when distinguishing between matched images) for alcohol, but not non-alcohol, images.

Methods

AAPS production details: Images were photographed in our laboratory using a Nikon D2× digital camera (Nikon Corp., Tokyo, Japan) with a Nikkor 60mm Micro f/2.8 lens (Nikon Corp., Tokyo, Japan). All images are color photographs with a resolution of 600 × 450 pixels (96 dpi, sRGB color format). Images were standardized on background color (white) and various modifiers and adapters were used to shape and control the light. Images were edited in Adobe Lightroom 6.3 (2015, Mountain View).

Participants: Participants consisted of registered users of Amazon Mechanical Turk (MTurk) via TurkPrime (www.turkprime.com). Participants on MTurk see a list of potential jobs (referred to as HITs), which include surveys, opinion polls, and psychological studies. MTurk has been extensively and reliably used by social psychologists to recruit representative American samples for research purposes (22,23). Participants were American citizens over 18 years of age, required to have consumed at least one alcoholic beverage in the past 30 days, and required to have previously completed over 100 HITs on MTurk with >95% approval rating. Participants provided e-consent and were compensated for their time and effort. The Committee on Human Research at the University of California, San Francisco approved all study protocols.

Procedures

Tasks were completed in the following order: (1) eligibility survey, (2) demographics questionnaire, (3) AUDIT, (4) Appetitive Rating Task, and (5) Arrow-Probe Matching Discrimination Task. All programming was done using Inquisit 4.0.8.0 (2014, Seattle). Participants were provided written instructions for all tasks and invited to complete as many practice trials for tasks 4 and 5 as they wished prior to performing the task. Since participants were completing our tasks remotely, four "attention check" questions were randomly inserted into the Appetitive Rating Task. To incentivize attention to the procedures, participants' HITs were not approved (and they were not compensated) if they got any of the "attention check" questions wrong or if they performed with <75% accuracy on the Arrow-Probe Task. This was based on pilot data of 40 participants, who were monitored in the laboratory while performing the Arrow-Probe Task, demonstrating accuracy ≥84% (M = 91.1%). Participants who successfully completed all tasks were invited to complete all tasks a second time two weeks later, which allowed for test-retest reliability analysis.

Alcohol use disorders identification test (AUDIT)

The AUDIT is a well-validated, 10-question, self-report questionnaire designed to identify people at risk of hazardous drinking over the past year. We defined "light drinkers" (LD) as participants with an AUDIT score ≤8 and "heavy drinkers" (HD) as participants with AUDIT scores >8 (24).

Appetitive rating task

Each AAPS image was displayed for three seconds, in a randomized order, on a white background in the center of the computer screen. Each displayed image was followed by two questions assessing incentive salience, or "wanting", and expected pleasure, or "liking", respectively (25). The questions "How much do you want the drink right now?" and "How much would you enjoy the drink right now?" were each followed by a visual analog scale (VAS). The VAS measured 80% of the screen width. The far left of the VAS was marked "I do not want the drink at all right now" for the "want" question and "I would not enjoy this drink at all right now" for the "like" question. The far right of the VAS was marked "I really want the drink right now" for the "want" question and "I would really enjoy this drink right now" for the "like" question. Participants moved a cursor along the VAS, and the rating was logged when the participant clicked a button on the bottom of the page to move to the next question. Scores were recorded as a number from 0 to 100. Participants completed one trial for each photograph, totaling 36 trials.

Arrow-probe matching discrimination task

Participants were instructed to focus on a fixation cross at the center of the screen. After 500ms, an arrow replaced the fixation cross, pointing either left or right. Concurrently, matched alcohol and non-alcohol image pairs appeared, one image on either side of the arrow. Image pairs appeared in a randomized order. The arrow and images were displayed for 200ms. In general, participants are unable to initiate an eye movement within 200 ms if they have to attend to a stimulus at the fixation location (the central arrow in our case) (26). Next, the words "alcohol" and "non-alcoholic drink" appeared, one on top of the other in a randomized configuration. Participants had five seconds to select the category they believed the arrow was pointing to in the preceding set using either the up or down arrow keys on the keyboard (see Figure 2). Participants were instructed to move as quickly as possible. Once a category was selected, or if no response was recorded after five seconds, the fixation cross appeared for another 500 ms followed by another arrow with the next pair of matched images. Participants completed 72 trials total, each pair appearing four times, twice with alcohol on the left (once with the arrow pointing left and once with the arrow pointing right) and twice with alcohol on the right (once with the arrow pointing left and once with the arrow pointing right). Accuracy and latency were recorded for each trial.

Data analysis

Data were inspected for normality both visually and using the Shapiro–Wilk Expanded Test. Data were assessed for outliers. One participant had an AUDIT score of 39, which was more than six standard deviations above the mean. Because our participants completed surveys anonymously via MTurk, we had no way of assessing whether this participant was misreporting and, thus, excluded their data from our analyses. Standard descriptive statistics were used to summarize the demographic data, appetitive ratings, and accuracy and latency in the arrow-probe matching discrimination task for each image category overall and for LD and HD separately. Homogeneity of variance was evaluated and assumptions were met as indicated by non-significant Levene's test. Thus, 2 (Beverage Category: Alcohol vs. Non-Alcohol; within subjects) × 2 (Group: LD vs. HD; between subjects factor) repeated measures ANOVA were performed for each outcome measure. Post-hoc tests were performed using paired and independent samples, two-tailed, Student's t-tests. We corrected for multiple comparisons using Bonferroni's Correction. To establish construct validity, we compared each participant's AUDIT score with their mean appetitive ratings ("want" and "like") and arrow-probe task accuracy and latency using Pearson's correlations. We used Pearson's correlations to determine test-retest reliability coefficients for appetitive ratings and arrow-probe latencies completed by the same participants two weeks apart.

Results

Participants

178 participants were included in analyses, 126 were defined as LD and 52 were defined as HD. See Table 1 for demographic characteristics. A total of 120 participants successfully repeated all tasks two weeks after initial completion and were included in test-retest reliability analysis.

AAPS descriptives

See Table 2 for image category characteristic details.

Overall "want" ratings of non-alcohol beverage images were significantly higher than those of alcohol beverages, $F(1, 176) = 34.38$, $p < 0.001$. Two-way interaction between Beverage Category and Group was significant, $F(1, 176) = 12.02$, $p = 0.001$. Post-hoc tests revealed that the effect of Beverage Category was only significant for the LD Group, $t(125) = -8.31$, $p < 0.001$, while not significant for the HD Group, $t(51) = -1.58$, $p = 0.12$. Furthermore, the HD Group rated alcohol images significantly higher than did the LD Group, $t(176) = -3.90$, $p < 0.001$, but there was no significant difference in "want" ratings for the non-alcohol images between the LD and HD Groups, $t(176) = 0.10$, $p = 0.92$.

Overall "like" ratings of non-alcohol beverage images were significantly higher than those of alcohol drinks, $F(1, 176) = 62.30$, $p < 0.001$. Two-way interaction between Beverage Category and Group was significant, $F(1, 176) = 14.26$, $p < 0.001$. Post-hoc tests revealed that the effect of Beverage Category was significant for both the LD Group, $t(125) = -10.89$, $p < 0.001$, and the HD Group, $t(51) = -2.40$, $p = 0.02$. Similar to "want" ratings, the HD Group reported significantly higher "like" ratings for alcohol images than did the LD Group, $t(176) = -5.02$, $p < 0.001$, but there was no significant difference for non-alcohol images between the LD and HD Groups, $t(176) = -1.11$, $p = 0.27$.

Alcohol beverage images were identified significantly more accurately overall in the arrow-probe task than non-alcohol beverage images, $F(1, 176) = 9.109$, $p = 0.003$. Two-way interaction between Beverage Category and Group was not significant,

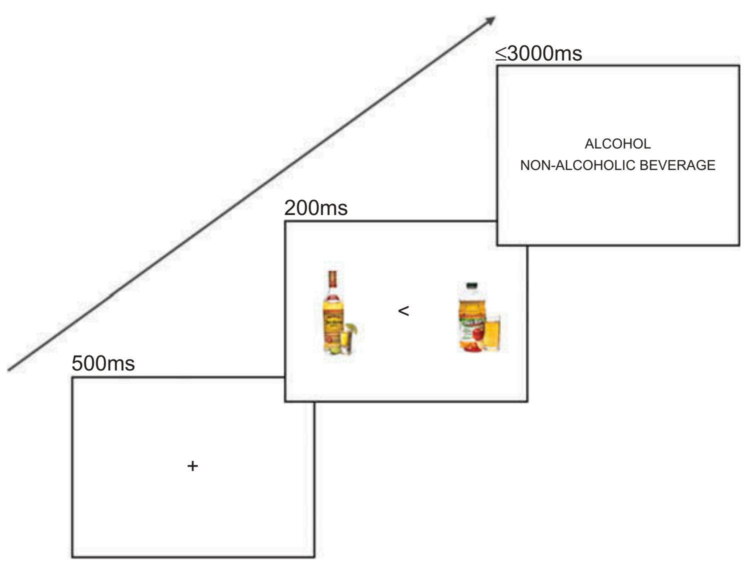

Figure 2. Schemata of the arrow-probe matching discrimination task. ms = milliseconds.

Table 1. Demographic characteristics.

	Total (N = 178)		LD (N = 126)		HD (N = 52)	
	Mean (SD)	Range	Mean (SD)	Range	Mean (SD)	Range
Age (years)	33.33(9.21)	19–59	34.31(9.42)	19–59	30.96(8.30)	21–54
Education (years)	14.79(2.02)	11–20	14.82(2.03)	12–20	14.71(2.01)	11–18
AUDIT score	6.67(5.33)	1–23	3.81(1.89)	1–8	13.60(4.57)	9–23
Gender	N	%	N	%	N	%
Female	95	53.37	74	58.73	21	40.38
Male	82	46.07	51	40.48	31	59.62
Genderqueer	1	0.56	1	0.79	0	0
Race						
AfAm/Black	8	4.49	5	3.97	3	5.77
Asian	7	3.93	4	3.17	3	5.77
White	153	85.96	111	88.10	42	80.77
More than one race	7	3.93	5	3.97	2	3.85
Other	3	1.69	1	0.79	2	3.85

LD = light drinkers (AUDIT≤8), HD = heavy drinkers (AUDIT>8), SD = standard deviation, AfAm = African-American, AUDIT = Alcohol Use Disorder Identification Test.

Table 2. AAPS image descriptives by category.

			Alcohol		Non-alcohol		
			Mean	SD	Mean	SD	p-value
Appetitiveness							
Want	*0–100*	Total	49.01	20.74	62.68	16.57	**<0.001**
		LD	45.13	20.40	62.72	16.40	**<0.001**
		HD	58.01	18.63	62.49	17.41	0.12
		LD:HD p-value	**<0.001**		0.92		
Like	*0–100*	Total	31.92	23.88	50.76	19.16	**<0.001**
		LD	26.31	22.09	50.34	18.80	**<0.001**
		HD	45.04	22.66	53.24	19.48	0.02
		LD:HD p-value	**<0.001**		0.27		
Arrow-Probe							
Accuracy	*%*	Total	91.67	6.50	90.07	6.49	**0.003**
		LD	91.99	6.16	90.67	6.39	<0.05
		HD	90.97	7.32	88.89	6.53	0.02
		LD:HD p-value	0.38		0.10		
Latency	*ms*	Total	1146.23	192.36	1144.43	183.43	0.18
		LD	1177.81	195.71	1155.26	184.52	0.02
		HD	1088.98	170.75	1126.70	182.31	**0.003**
		LD:HD p-value	0.01		0.41		

SD = standard deviation, LD = light drinkers, HD = heavy drinkers, ms = milliseconds, bold values survived Bonferroni correction for multiple comparisons.

$F(1, 176) = 0.364$, $p = 0.55$. As hypothesized, mean accuracy was above 89.9% for both categories overall; however, the HD Group was only 88.89% accurate in recognizing non-alcohol.

Overall, there was no significant difference between the two Beverage Categories in latency in the arrow-probe task, $F(1, 176) = 1.85$, $p = 0.18$. Two-way interaction between Beverage Category and Group was significant, $F(1, 176) = 15.04$, $p < 0.001$. Post-hoc tests revealed that the LD Group was significantly faster in correctly recognizing non-alcohol images versus alcohol images, $t(125) = 2.34$, $p = 0.02$. Conversely, the HD Group was significantly faster in correctly recognizing alcohol versus non-alcohol images, $t(51) = -3.09$, $p = 0.003$. Furthermore, the HD Group was significantly faster at recognizing alcohol than the LD Group, $t(176) = 2.59$, $p = 0.01$, but there was no significant difference in arrow-probe reaction time for non-alcohol images between the LD and HD Groups, $t(176) = 0.83$, $p = 0.41$.

Construct validity

Participants' AUDIT scores ($M = 6.67$, $SD = 5.33$) correlated significantly with alcohol "want", $r(176) = 0.27$, $p < 0.001$, and "like", $r(176) = 0.36$, $p < 0.001$, ratings, but not with non-alcohol "want", $r(176) = -0.03$, $p = 0.71$, and "like", $r(176) = 0.04$, $p = 0.63$, ratings. AUDIT scores were also correlated with (alcohol – non-alcohol) difference scores for "want", $r(176) = 0.26$, $p < 0.001$, and "like", $r(176) = 0.31$, $p < 0.001$, ratings. AUDIT scores were not correlated with arrow-probe accuracy for alcohol, $r(176) = -0.02$, $p = 0.82$, or non-alcohol, $r(176) = -0.06$, $p = 0.45$. AUDIT scores negatively correlated with latency in the arrow-probe task (i.e., higher AUDIT scores correlated with faster reaction times) significantly for alcohol images, $r(176) = -0.22$, $p = 0.004$, but not for non-alcohol images, $r(176) = -0.10$, $p = 0.18$. See Figure 3.

Test-retest reliability

Test-retest reliability coefficients for alcohol and non-alcohol, respectively, were as follows: "want" ratings 0.80 and 0.77, "like" ratings 0.72 and 0.73, arrow-probe latency 0.83 and 0.75.

Discussion

We produced a set of 36 alcohol and non-alcohol beverage photographs standardized for physical image and contextual characteristics. We validated the images in two ways: (1) face validity – alcohol and non-alcohol beverages matched for general color, shape, and complexity were adequately distinguishable when flashed briefly (200 ms) in peripheral vision, and (2) construct validity – variations in AUDIT scores were associated with variations in appetitive ratings and performance on our arrow-probe task designed to test implicit attentional bias. Image appetitive ratings and latency in the arrow-probe task were reliable when assessed at two timepoints separated by two weeks. There are no previously published, validated, reliable, American-relevant sets of alcohol images created in a standardized fashion.

Overall, participants rated non-alcohol images as more appetitive than alcohol images; however, this is consistent with Pronk et al. (13), who showed a similar trend for self-reported urge to drink after viewing images of alcohol versus non-alcohol. Also consistent with Pronk et al. (13), our alcohol images – but not non-alcohol images – received significantly higher appetitiveness ratings and resulted in shorter latency in our arrow-probe task when assessed by HD versus LD, and there were no differences in appetitive ratings or arrow-probe latencies for non-alcohol images between LD and HD. Furthermore, in line with our hypotheses, AUDIT scores significantly predicted appetitive ratings and implicit attentional bias for alcohol but not for non-alcohol images; however, the magnitude of these correlations were relatively modest. To further prove the hypothesis that lower appetitive ratings of alcohol versus non-alcohol images were driven by those participants who consume less alcohol in general (i.e., LD), we demonstrated that appetitive rating difference (alcohol – non-alcohol) scores correlated significantly with AUDIT scores. Moreover, based on mean AUDIT scores, our "HD" subsample did not objectively drink very "heavily" relative to the possible range of AUDIT scores, HD in our study being ranked risk level II of IV by the World Health Organization (i.e., suggested intervention is "simple advice") (24). We conclude that AAPS provides valid alcohol stimuli for both explicit and implicit tests of cue reactivity. Given the assumption that alcohol has higher appetitive appeal for HD than for LD (27), variation in the pictures (alcohol versus non-alcohol) is associated with variation in self-reported appetitive responses and implicit bias for alcohol pictures, but not non-alcohol pictures, in HD versus LD.

Several limitations exist when considering the validity and relevance of this image set. Importantly, AAPS still requires validation for certain patient populations, such as those with confirmed alcohol use disorder. We are currently using AAPS images in an Approach-Avoidance Task for patients with moderate-severe alcohol use disorder and post-traumatic stress disorder. Secondly, we collected no data on brand recognition or alcohol preference from our participants. Furthermore, it is uncertain if face

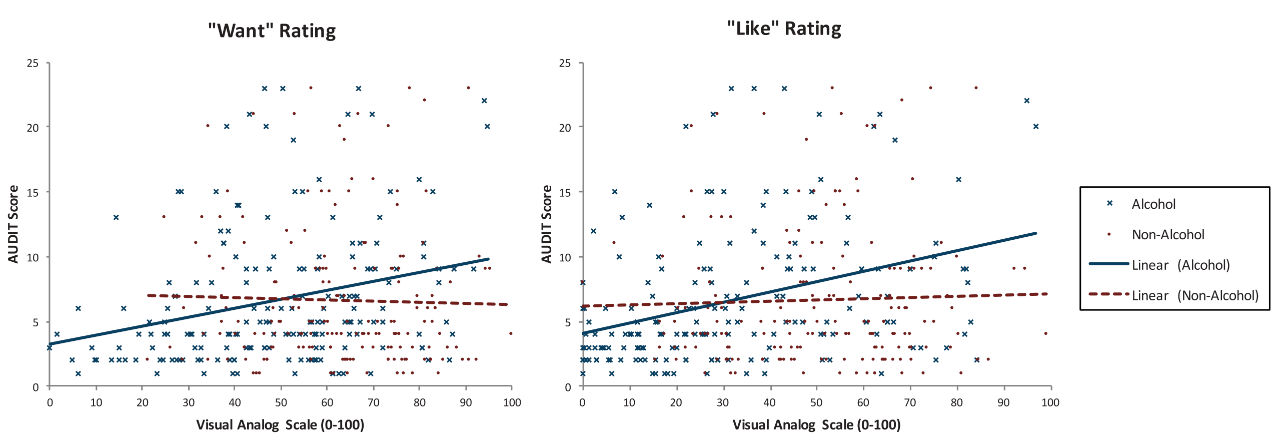

Figure 3. Pearson's correlations of AUDIT scores with appetitive ratings. "Want": Alcohol $r^2 = 0.07$, Non-Alc $r^2 = 0.0008$. "Like": Alcohol $r^2 = 0.13$, Non-Alc $r^2 = 0.001$.

validity for AAPS will generalize to a broader American population. However, we selected for a variety of alcohol types (beer = 7, wine = 3, spirits = 8), and AAPS as a whole was created so that distinct subsets of images would appeal to a range of demographic populations. We provide summary data for each individual image in our Supplemental Material. Lastly, it is difficult to conclude that AAPS represents an improvement above *ad-hoc* beverage image sets, because these stimuli generally still show appetitive ratings and attentional bias for alcohol pictures. Future comparison studies would be helpful to address this question.

In conclusion, cue reactivity paradigms to date have been highly heterogeneous. However, as technology advances and data sharing becomes increasingly convenient the bar for standardization of stimuli is being raised. The use of standardized images across studies and paradigms can lead to more reproducible and comparable datasets. In order to fill a gap in the literature, we are making this American-relevant image set of matched alcohol and non-alcohol beverages available to the scientific community for the research and treatment of alcohol use disorder. **For complimentary access to AAPS, interested researchers can contact our lab at BandLab@ucsf.edu**.

Acknowledgements

We would like to thank Ryan Darcy, executive producer at Castile Production, and his creative team for donating their time and talent in the name of science. We would like to also acknowledge the support of the Department of Defense and the UCSF Institute for Translational Neuroscience.

Declaration of interest

The authors report no relevant financial conflicts.

ORCID

Christopher S. Stauffer http://orcid.org/0000-0002-0888-095X

References

1. Naqvi NH, Morgenstern J. Cognitive neuroscience approaches to understanding behavior change in alcohol use disorder treatments. Alcohol Res Curr Rev 2015;37:29.
2. Reynolds EK, Monti PM. In the Wiley-Blackwell handbook of addiction psychopharmacology. West Sussex, UK: Wiley-Blackwell; 2013, 381–410.
3. Stormark KM, Torkildsen Ø. Selective processing of linguistic and pictorial food stimuli in females with anorexia and bulimia nervosa. Eating Behav 2004;5:27–33.
4. Boswell RG, Kober H. Food cue reactivity and craving predict eating and weight gain: a meta-analytic review. Obesity Rev 2016;17:159–177.
5. Ihssen N, Cox WM, Wiggett A, Fadardi JS, Linden DE. Differentiating heavy from light drinkers by neural responses to visual alcohol cues and other motivational stimuli. Cereb Cortex 2011;21:1408–1415.
6. Palfai TP, Kantner CK, Tahaney KD. The image-based alcohol-action implicit association test. J Behav Ther Exp Psychiatry 2016;50:135–138.
7. Grüsser SM, et al. Cue-induced activation of the striatum and medial prefrontal cortex is associated with subsequent relapse in abstinent alcoholics. Psychopharmacology 2004;175:296–302.
8. Naqvi NH, et al. Cognitive regulation of craving in alcohol-dependent and social drinkers. Alcohol Clin Exp Res 2015;39:343–349. doi:10.1111/acer.12637.
9. Myrick H, et al. Effect of naltrexone and ondansetron on alcohol cue–induced activation of the ventral striatum in alcohol-dependent people. Arch Gen Psychiatry 2008;65:466–475.
10. Mann K, Hermann D. Individualised treatment in alcohol-dependent patients. Eur Arch Psychiatry Clin Neurosci 2010;260:S116–S120. doi:10.1007/s00406-010-0153-7.
11. Miller MA, Fillmore MT. The effect of image complexity on attentional bias towards alcohol-related images in adult drinkers. Addiction 2010;105:883–890. doi:10.1111/j.1360-0443.2009.02860.x.
12. Lang PJ, Bradley MM, Cuthbert BN. In Technical Report A-8. University of Florida, Gainesville, FL, 2008.
13. Pronk T, van Deursen DS, Beraha EM, Larsen H, Wiers RW. Validation of the Amsterdam beverage picture set: a controlled picture set for cognitive bias measurement and modification paradigms. Alcohol Clin Exp Res 2015;39:2047–2055. doi:10.1111/acer.12853.
14. Pulido C, Brown SA, Cummins K, Paulus MP, Tapert SF. Alcohol cue reactivity task development. Addict Behav 2010;35:84–90.
15. Blechert J, Meule A, Busch NA, Ohla K. Food-pics: an image database for experimental research on eating and appetite. Front Psychol 2014;5:617. doi:10.3389/fpsyg.2014.00617.
16. Stacy AW, Wiers RW. Implicit cognition and addiction: a tool for explaining paradoxical behavior. Ann Rev Clin Psychol 2010;6:551.
17. Palfai TP, Kantner CK, Tahaney KD. The image-based alcohol-action implicit association test. J Behav Ther Exp Psychiatry 2016;50:135–138. doi:10.1016/j.jbtep.2015.07.002.
18. Lee E, Namkoong K, Lee CH, An SK, Lee BO. Differences of photographs inducing craving between alcoholics and non-alcoholics. Yonsei Med J 2006;47:491–497.
19. Nees F, Diener C, Smolka MN, Flor H. The role of context in the processing of alcohol-relevant cues. Addict Biol 2012;17:441–451. doi:10.1111/j.1369-1600.2011.00347.x.

20. Verma A, Brysbaert M. A validated set of tool pictures with matched objects and non-objects for laterality research. Laterality 2015;20:22–48. doi:10.1080/1357650X.2014.914949.
21. Borsboom D, Mellenbergh GJ, van Heerden J. The concept of validity. Psychological Review. 2004;111 (4):1061–1071.
22. Litman L, Robinson J, Rosenzweig C. The relationship between motivation, monetary compensation, and data quality among US- and India-based workers on mechanical turk. Behav Res Methods 2015;47:519–528.
23. Buhrmester M, Kwang T, Gosling SD. Amazon's mechanical turk: a new source of inexpensive, yet high-quality, data? Perspect Psychol Sci J Assoc Psychol Sci 2011;6:3–5. doi:10.1177/1745691610393980.
24. Babor TF, Higgins-Biddle JC, Monteiro MG. The alcohol use disorders identification test (AUDIT): guidelines for use in primary care, 2nd edn. WHO/MSD/MSB/01.6a. Geneva, Switzerland: World Health Organization, Department of Mental Health & Substance Dependence; 2001.
25. Robinson TE, Berridge KC. Incentive-sensitization and addiction. Addiction 2001;96:103–114.
26. McSorley E, Haggard P, Walker R. Time course of oculomotor inhibition revealed by saccade trajectory modulation. J Neurophysiol 2006;96:1420–1424.
27. Townshend J, Duka T. Attentional bias associated with alcohol cues: differences between heavy and occasional social drinkers. Psychopharmacology 2001;157:67–74.