The effect of question structure on self-reported drinking:

Ascending versus descending order effects

Sean P. Mackinnon

Dalhousie University

Samantha Firth

Dalhousie University

Author Note

Sean P. Mackinnon, Department of Psychology and Neuroscience, Dalhousie University, Halifax, Nova Scotia, Canada. Samantha Firth, Department of Psychology and Neuroscience, Dalhousie University, Halifax, Nova Scotia, Canada.

Correspondence concerning this article should be addressed to Sean P. Mackinnon, Department of Psychology and Neuroscience, Dalhousie University, Life Sciences Centre, 1355 Oxford Street, PO Box 15000, Halifax, Nova, Scotia, Canada, B3H 4R2. Email: mackinnon.sean@dal.ca, Phone: (902) 494-3385, Fax: (902) 494-6585.

Acknowledgements

This work was supported by the Dalhousie University Health Sciences SSHRC Research Grant [#36317], Halifax, Nova Scotia.

The authors thank Lucas Nystrom, Catherine Robertson, Makaila Koshurba and Lee Cai Pook-Clendenning for their research assistance.

Sean Mackinnon created the study design, worked on data analysis and manuscript writing, and supervised data collection. Samantha Firth handled contact with participants and data collection, and assisted with data analysis and manuscript writing.

Highlights

- Primacy effects should occur in measures related to alcohol use
- These effects will occur most when responses are presented in descending order
- Primacy effects were seen for sensation-seeking and enhancement motives
- Participants endorsed “Strongly Agree” more often in the descending condition

Abstract

Response-order effects refer to differences in participant responding based on the order of questionnaire response options. Theory predicts primacy effects (options presented earlier are more frequently endorsed). We predicted primacy effects for unipolar scales related to alcohol. Undergraduates (N = 791) completed questionnaires on drinking motives, personality, and alcohol problems. Participants were randomly assigned to ascending (highest-to-lowest) or descending (lowest-to-highest) response-order conditions. Pre-registered analyses found primacy effects for sensation-seeking and enhancement motives. There was no moderation by sex, ethnicity, or age. Experimental condition had no impact on concurrent validity. Exploratory analyses found that undergraduates selected “Strongly Agree” more often in the descending condition and primacy effects for alcohol problems. Overall, we found evidence for small primacy effects in undergraduate students.

Keywords: alcohol; drinking motives; personality; primacy effect; response-order effects

1. Introduction

Responses on self-report questionnaires can be strongly influenced by features of the research instrument, such as reference periods and juxtaposition with other measures (Schwarz, 1999). In a review of the literature, Krosnick & Presser (2009) found that visual presentation of unipolar and categorical questionnaire items typically results in primacy effects (i.e., placement of responses at the beginning of a list increases the probability they will be chosen). In contrast, they found that oral presentation tends to result in recency effects (i.e., placement at the end of a list increases the probability that option will be chosen). In the present study, we examine response-order effects in visually-presented items when measuring alcohol problems and related personality measures. That is, whether primacy effects exist for unipolar self-report scales when using a between-subjects experiment comparing ascending (i.e., starting with strongly disagree / lowest frequency of behavior) to descending (i.e., starting with strongly agree / highest frequency of behavior) order of presentation.

1.1. Primacy effects

Krosnick and Alwin (1987) propose that questionnaire responses will demonstrate a primacy effect. Early items are thought to establish a cognitive framework that guides interpretation of later items and are subjected to deeper cognitive processing by virtue of being presented first. Moreover, people tend to choose the first acceptable option in a list (rather than the most accurate or optimal choice) as a way to minimize cognitive effort (i.e., "satisficing"). Hogarth and Einhorn's (1992) belief-adjustment model explains response-order effects via anchoring and adjustment. They argue that response-order effects depend on response mode, task length, and information complexity. They propose two response modes: (a) Step-by-Step, where participants make new judgements after each response option is presented; and (b) End-of-Sequence, where participants make a single net judgement after all response options are presented. The model proposes that End-of-Sequence processing will result in primacy effects, while Step-by-Step processing will result in recency effects, with task complexity and length operating as moderators. In the present study, participants are using end-of-sequence processing with a short, simple task. Thus, the belief-adjustment model predicts primacy effects.

Generally speaking, primacy effects due to response order have been found in numerous domains, including self-reported health (Garbarski, Schaeffer, & Dykema, 2015), subjective importance of job characteristics (Krebs & Hoffmyer-Zlotnik, 2010), and personality traits

(Krosnick & Alwin, 1987). Malhortra (2008) found that participants were more likely to select the first two options on unipolar rating scales in a descending condition. Moreover, they found that primacy effects were more pronounced when participants completed surveys quickly, and when participants had lower education levels. Overall, there is strong evidence for primacy effects for questionnaires using unipolar scales.

1.2. Tests of concurrent validity

Response-order effects are a form of method variance, which introduces measurement error. If a relationship actually exists in the population, measurement error reduces the magnitude of correlations (Podsakoff, Mackenzie, Lee, & Podsakoff, 2003). Thus, it is possible to test the impact of methodological changes on concurrent validity by examining established, theory-driven correlations. That is, does the magnitude of well-established relationships between variables differ depending on whether items are presented in ascending vs. descending order? For example, perhaps participants are biased to select "Strongly Agree" when it is presented first (ascending), but are not similarly biased to select the first option when "Strongly Disagree" is presented first (descending). If this were the case, we might expect larger correlations between theoretically-linked variables in the descending condition. In contrast, if responding in both conditions are equally biased, we might see a primacy effect (i.e., larger means in the descending condition), but no difference in the magnitude of the correlations between variables.

We assessed concurrent validity by examining the well-established, positive relationships between personality, drinking motives, and alcohol problems. We used a 4-factor model of personality (Woicik, Stewart, Pihl, & Conrod, 2009) which focuses on traits that confer vulnerability to alcohol misuse. Specifically, we used the Substance Use Profile Scale (Woicik et al., 2009), which measures impulsivity (lack of premeditation), sensation seeking (need for intense, novel experiences), hopelessness (bleak expectations about oneself and the future), and anxiety sensitivity (fear of anxiety-related sensations). We also used a common model of drinking motivation (Cooper, 1994) which includes four dimensions: Social (drinking to increase social facilitation), enhancement (drinking to enhance positive emotions), conformity (drinking to avoid social rejection), and coping motives (drinking to reduce positive emotions). Though some relationships are larger than others, these measures are usually positively correlated with alcohol problems and each other (Mackinnon, Kehayes, Clark, Sherry, & Stewart, 2014).

1.3. Current Study

We used a two-condition between-subjects design to examine response-order effects in the area of personality, drinking motives, and alcohol problems. Though various studies have examined response-order effects (Krosnick & Presser, 2009), few studies have examined this in the area of alcohol problems, leading some to call for more research in this area (Dawson & Room, 2000). Moreover, though prior research has generally established the existence of primacy effects when examining means (e.g., Krebs & Hoffmyer-Zlotnik, 2010), they have not typically been able to distinguish what effect this has on validity (i.e., if one method is preferable over the others). Finally, prior research has not tended to incorporate common demographic covariates and moderators such as age, sex and ethnicity. Given that these variables are associated with alcohol problems (Wilsnack, Wilsnack, Kristjanson, Vogeltanz-Holm, & Gmel, 2009), it seems prudent to examine whether they influence the magnitude of primacy effects due to response order on such questionnaires. Finally, few researchers in this area share data in an open-access database and pre-register hypotheses (c.f., Garbarski et al., 2015). In contrast, the present study uses pre-registered hypotheses, open-access data, and open materials[1].

Given the expectation that presenting questionnaire items in descending order will result in primacy effects, our primary hypothesis was:

> H1: Participants will endorse higher levels of alcohol misuse, risky personality traits, and drinking motivations, when closed-ended responses are presented in descending order, as opposed to being presented in ascending order.

We had two secondary research questions. These analyses are important to test for concurrent validity and potential demographic moderators. However, prior research was insufficient to suggest a direction for effects a-priori; thus, we present these questions as non-directional research questions.

> RQ1: Does the effect specified in H1 vary across various subgroups? Specifically, is this effect moderated by sex, age, or ethnicity?
>
> RQ2: Do the correlational relationships between these three variables (i.e., personality, drinking motives, and alcohol problems) vary depending on whether or not responses are presented in descending order, as opposed to being presented in ascending order? That is, does experimental condition moderate these well-established relationships?

---

[1]A time-stamped pre-registration of hypotheses can be found here https://osf.io/prvu5/. All study materials, raw data, and syntax are located at https://osf.io/9nq3c/.

2. Method

2.1. Power Analysis

Power analysis was conducted for the primary hypothesis and outcome (i.e., the relationship between experimental condition and AUDIT-C scores) for a Poisson regression using G*Power software. Assuming a small effect size (odds ratio = 1.10; or a 10% increase in AUDIT-C scores), α of .05, and a mean score for the AUDIT-C of 4.17 (Bradley et al., 2007), we required 791 participants to achieve 80% power.

2.2. Participants

A total of 791 participants were recruited through online ads (n = 95) and the Psychology Department Participant Pool (n = 696). Participants were primarily young (M_{age} = 21.08, SD_{age} = 5.87), female (80.4%), and Caucasian (74.6%). There were no inclusion criteria other than a requirement for participants to reside in the Halifax Regional Municipality, as we were also interested in whether or not the experimental manipulation affected the number of abstainers (i.e., participants who reported not drinking).

2.3. Measures

2.3.1. Demographics

The demographic questionnaire was composed of 3 items and was used to obtain participants' age, sex, and ethnicity. Age was measured in years. Sex was dichotomized as male/female, with participants indicating "other" (n = 3) omitted from analyses. Participants reported race using open-ended text; this variable was dichotomized into Caucasian/non-Caucasian for analyses by the second author.

2.3.2. Alcohol Problems

The short form of the Alcohol Use Disorders Identification Test (AUDIT-C) is a 3-item questionnaire that measured participants' frequency and quantity of alcohol consumption and frequency of heavy episodic drinking over the past year. Alcoholic drinks were defined as a 12-ounce can, glass or bottle of beer or cooler, a 4-ounce glass of wine, or a drink containing 1 shot of liquor/spirits[2]. Each item was rated using 5-point scale. Bush, Kivlahan, McDonell, Fihn, &

[2]Due to a typographical error, all participants saw instructions reading "4 oz glass of wine" rather than the intended "5oz." This is unlikely to substantially alter the results.

Bradley (1998) found that the AUDIT-C has excellent sensitivity and specificity when predicting alcohol abuse/dependence, and is highly similar to the full AUDIT.

2.3.3. Drinking Motives

Drinking motives were measured using the Drinking Motives Revised – Short Form questionnaire (DMQ-R-SF; Kuntsche & Kuntsche, 2009). This measure includes 4 subscales: social, enhancement, coping, and conformity. Participants indicated how often they drank for each reason in the past 12 months using a 4-point Likert scale. Prior research shows this measure has good internal consistency ($\alpha s \geq .70$) and factorial validity (Kuntsche & Kuntsche, 2009).

2.3.4. Personality

Personality was measured using the Substance Use Risk Profile Scale (SURPS, Woicik et al., 2009). This scale measures four personality traits related to alcohol outcomes and motives (i.e., impulsivity, sensation seeking, hopelessness, and anxiety sensitivity). Participants responded to items pertaining to how they may have felt or behaved over the past several years using a 4-point Likert scale. Prior research shows that this measure has good internal consistency ($\alpha s \geq .70$) and construct validity, with all subscales showing strong relationships with the original measures they were derived from (Woicik et al., 2009).

2.4. Procedure

Participants signed up for the study via the online participant pool or by emailing the second author, who provided participants with a link to the online questionnaire using Opinio software (Object Planet, 2016). We used a between-subjects experimental design with two conditions. Participants were randomly assigned to one condition, with 385 (48.7%) in the ascending condition and 406 (51.3%) in the descending condition. In both conditions, participants received a baseline questionnaire along with the AUDIT-C (Bradley et al., 2007), the 12-item DMQ-R-SF (Kuntsche & Kuntsche, 2009), and the SURPS (Woicik et al., 2009). In the ascending condition, participants viewed closed-ended response options to the questionnaire items from lowest to highest (e.g., strongly disagree to strongly agree). In the descending condition, participants viewed response options from highest to lowest (e.g., strongly agree to strongly disagree). Skip logic was implemented for the DMQ-R-SF; participants did not complete this questionnaire if they reported never drinking alcohol on the AUDIT-C. All questionnaires were completed online on participants' own computers. After completing the surveys, participants were debriefed and had the option to leave their email address in order to be

entered into a draw to win a $100 Amazon gift card. Eligible psychology students also received 0.5 bonus credits.

2.5. Data Analytic Strategy

Hypothesis 1 and RQ1 were tested using generalized linear models (GLiM) in SPSS 23 software with robust estimates of standard errors. GLiM is more flexible and requires less stringent assumptions than ANOVA-based methods. We compared Normal, Poisson, and negative binomial regression models for each outcome using nested model comparison. Models with the lowest BIC values were preferred. The primary hypothesis (H1) used experimental condition (ascending vs. descending) as the predictor, and each of the 9 subscales from our measures (i.e., alcohol problems, motives, personality) as dependent variables in separate analyses. For RQ1, we added age, sex, and ethnicity as predictors to the regression models described in H1. In this analysis, we also added multiplicative interaction terms to the model (i.e., all possible two-way interactions with condition) using mean-centering to for age and effect coding for dichotomous predictors.

To test RQ2, we calculated Spearman rank correlations for all possible combinations of the 9 variables in our study, within each condition separately (i.e., 36 correlations per condition for a total of 72). Next, we calculated a weighted average of all 36 correlation coefficients for each condition, and then compared the magnitude of that average correlation (i.e., tested the null hypothesis that $r_{ascending} - r_{descending} = 0$) using a random-effects meta-analysis in SPSS. This single overall meta-analytic test served as the overall test of RQ2. Reducing the number of tests conducted helps combat the familywise error-rate problem. Because non-drinkers did not complete the DMQ-R-SF, this analysis is conducted on the subset of participants who drank alcohol in the past year.

2.6. Changes from Pre-registration

A few errors in the pre-registered analysis plan were corrected prior to analysis. First, we used sums rather than averages for all subscale totals because Poisson models require integers. Second, our pre-registered plan said there are 45 correlations per condition; however, with 9 variables, there are only 36 per condition. Finally, we used effect coding (-1, 1) rather than dummy coding (0, 1) for categorical variables. Though this has no impact on testing H1, effect coding is required for classical interpretation of effects in the presence of interaction terms when testing RQ1 (Kugler, Trail, Dziak, & Collins, 2012).

3. Results

3.1. Hypothesis test

Distributions for Generalized Linear Models were selected using ΔBIC values, with lower values being preferred (Supplementary Table 1). Poisson models were preferred for coping, conformity, and hopelessness; normal models were preferred for all other outcomes. Results for hypothesis tests are presented in Table 1. Overall, hypotheses were supported for 2 of 9 outcome variables. Specifically, means were lower in the descending condition for enhancement motives and sensation-seeking. Though results trended towards primacy effects for alcohol problems, hypotheses were not supported for all other variables.

3.2. Moderation

Models with age, sex, race, and their two-way interactions with condition were run to test RQ1 (Table 2). Once covariates were added to the model, the effect of experimental condition remained statistically significant only for sensation seeking. Age was associated with lower social and conformity motives. Being Caucasian was associated with higher alcohol problems, social motives, and enhancement motives and lower hopelessness and impulsivity. Men reported higher levels of alcohol problems, social motives, enhancement, conformity, sensation seeking, and impulsivity, while women reported higher levels of anxiety sensitivity. Only 1 of 27 interaction effects was statistically significant (i.e., the effect of condition on social motives was larger for Caucasian participants than for non-Caucasian participants). Data are visualized using violin & jitter plots in supplementary figures 1-4.

3.3. Meta-analysis

When all 72 pairwise spearman correlations were meta-analyzed, there was substantial heterogeneity, $\chi^2(71) = 590.20$, $p < .001$, suggesting a random-effects model was justifiable. Overall, the average correlation in the ascending condition ($r = .16$) did not differ from the average correlation in the descending condition ($r = .17$), $\chi^2(1) = 0.07$, $p = .80$. Moreover, none of the 36 pairwise comparisons were statistically significant in exploratory analyses (Table 3). Thus, experimental condition did not appear to affect concurrent validity. Scatterplots of all relationships split by condition with loess regression lines are presented in supplementary figures 5-8.

3.4. *Exploratory non-preregistered analysis.* Data were explored at the item-level to see if certain response options were endorsed more than others. Contingency tables presenting the

percentage of endorsement of each response option split by condition for all 38 items are presented in Supplementary Table 2. The AUDIT-C did not show a consistent pattern across all three items. The remaining items showed a weak pattern wherein the "Strongly Agree" option was endorsed slightly more often in the descending condition. As a more powerful test of this observation, we converted the data to long format and ran a generalized estimating equation with items nested within people. The 4-point Likert scale data were treated as ordinal (ordinal logistic regression), condition was entered as a categorical predictor, and item (i.e., a 35-category variable indicating item wording) was added as a covariate. There was a main effect of condition, $\chi^2(1) = 5.46$, $p = .019$, and item, $\chi^2(34) = 3225.94$, $p < .001$. The effect was driven primarily by people in the descending condition endorsing the "Strongly Agree" option more frequently than in the ascending condition (55.3% vs. 44.7%). Rates of endorsement were comparatively equal for the Slightly Agree/Agree (50.5% vs. 49.5%), Slightly Disagree/Disagree (50.7% vs. 49.3%), and Strongly Disagree (50.1% vs. 49.9%) response options. Bar plots displaying these results and results split by age, sex, and race are located in supplementary figures 9-12.

For the AUDIT-C, we also explored a zero-inflated negative binomial regression analysis with a logit link using the STAT ZEROINF extension in SPSS. This model fit the data better than the normal model, $\Delta BIC = 56$. Here, there was an effect of condition when predicting the count portion of the data, $B = 0.10$, $SE = 0.04$, $p = .02$. However, condition had no effect on the zero-inflated portion of the data, $B = 0.16$, $SE = 0.28$, $p = .56$[3]. Overall, this exploratory analysis suggested that non-abstainers in the descending condition reported more alcohol problems than people in the ascending condition.

We also re-analyzed the data in Table 3 using unstandardized coefficients from linear regression. Results were largely unchanged. Results of this analysis can be found in supplementary table 3.

4. Discussion

Overall, there was some support for the effect of question order on alcohol-related outcome variables. Means were slightly higher in the descending condition than in the ascending condition only for sensation-seeking and enhancement motives. Of these two effects, only sensation-seeking remained statistically significant once controlling for age, sex, and race.

[3]The effect of condition remained significant when controlling for sex and race.

Moreover, there were no interactions between experimental condition and these demographic variables. Finally, experimental condition did not appear to affect the magnitude of the correlations between outcome variables, suggesting that the order of response presentation does not greatly impact the concurrent validity of these measures. Though pre-registered analyses produced primacy effects for only 2 of 9 outcomes, exploratory analyses suggested that participants were more likely to endorse "Strongly Agree" in the descending condition, and that primacy effects were found for alcohol problems only for non-abstainers (i.e., the count portion of a zero-inflated model).

Krosnick and Alwin (1987) proposed primacy effects exist because people tend to choose the first acceptable option in a list to minimize cognitive effort. Consistent with this notion, exploratory analyses suggested participants were more likely to endorse "Strongly Agree" in the descending condition. The strongest effect was for sensation seeking. People high in sensation seeking are approach-oriented, and are motivated by the need for exciting and novel experiences (Mackinnon et al., 2014). Therefore, item content for sensation seeking asks about intense, but relatively uncommon sensory experiences (e.g., "I would like to skydive"). Thus, primacy effects might emerge more strongly for sensation-seeking because most participants have not previously thought about such activities, and are thus more likely to take cognitive shortcuts. Moreover, it is notable that the SURPS questionnaire was the last (and longest) questionnaire administered, which might also affect the desire to minimize cognitive effort. Krosnick and Presser (2009) suggest three possible solutions for circumventing response-order effects in personality scale measurement: counterbalancing presentation order, keeping questionnaires short and simple, and testing accountability by occasionally asking participants to justify their answers.

The present study has numerous limitations. It utilized a homogenous university student sample, with comparatively little variance in age, sex, and race making it difficult to test for interaction effects on these variables. Thus, results cannot be generalized to a population beyond university students and interaction effects are likely underpowered. The questionnaire we administered was short, only collecting measures reported in this paper. Research suggests that primacy effects would emerge more strongly when participants are fatigued (Krosnick & Presser, 2009). Anecdotally, many (perhaps most) studies in personality administer a large battery of questionnaires to participants; thus, the present study may have underestimated the magnitude of

the effect in more typical testing circumstances. Finally, 11.8% of participants (n = 93) did not drink alcohol in the past 12 months. This reduced the sample size for analyses on drinking motives, which in turn reduced statistical power. Thus, analysis on these variables may have been underpowered.

4.1. Conclusion

There is a large literature suggesting that question order has a small influence on questionnaire responding in a variety of domains (e.g., Garbarski et al., 2015; Krebs & Hoffmyer-Zlotnik, 2010). Nonetheless, it does appear that the magnitude of the effect in this domain (i.e., personality and alcohol use) is small enough to be of little practical concern for most purposes. Response option order did not appear to affect the correlational relationships, which is the target of hypotheses for most personality research. However, confidence interval widths were wide for these tests (about ±.30), so small effects on concurrent validity may still exist. When the purpose of research is to accurately estimate the mean levels of constructs, or when studied effects are very small, it is likely prudent to report this methodological detail to facilitate interpretation of results, or to use strategies aimed at reducing response-order biases.

References

Bradley, K. A., DeBenedetti, A. F., Volk, R. J., Williams, E. C., Frank, D., & Kivlahan, D. R. (2007). AUDIT-C as a brief screen for alcohol misuse in primary care. *Alcoholism: Clinical and Experimental Research, 31,* 1208-1217.

Bush, K., Kivlahan, D. R., McDonell, M. B., Fihn, S. D., & Bradley, K. A. (1998). The AUDIT alcohol consumption questions (AUDIT-C): an effective brief screening test for problem drinking. *Archives of internal medicine*, *158*, 1789-1795.

Cooper, M. L. (1994). Motivations for alcohol use among adolescents: Development and validation of a four-factor model. *Psychological Assessment, 6*, 117-128.

Dawson, D. A., & Room, R. (2000). Towards agreement on ways to measure and report drinking patterns and alcohol-related problems in adult general population surveys: The Skarpö Conference overview. *Journal of Substance Abuse, 12,* 1-21.

Garbarski, D., Schaeffer, N. C., & Dykema, J. (2015). The effects of response option order and question order on self-rated health. *Quality of Life Research*, *24*, 1443-1453.

Hogarth, R. M., & Einhorn, H. J. (1992). Order effects in belief updating: The belief-adjustment model. *Cognitive psychology*, *24*(1), 1-55.

Krebs, D., & Hoffmeyer-Zlotnik, J. H. (2010). Positive first or negative first? Effects of the order of answering categories on response behavior. *Methodology, 6,* 118-127.

Krosnick, J. A., & Alwin, D. F. (1987). An evaluation of a cognitive theory of response-order effects in survey measurement. *Public Opinion Quarterly*, *51*, 201-219.

Krosnick, J.A., & Presser, S. (2009). Question and Questionnaire Design. *Handbook of Survey Research, 2*(3), 263-314.

Kugler, K. C., Trail, J. B., Dziak, J. J., & Collins, L. M. (2012). Effect coding versus dummy coding in analysis of data from factorial experiments. *The Methodology Center, The Pennsylvania State University*. Retrieved July 18, 2017 from https://methodology.psu.edu/media/techreports/12-120.pdf

Kuntsche, E., & Kuntsche, S. (2009). Development and validation of the drinking motive questionnaire revised short form (DMQ-R SF). *Journal of Clinical Child & Adolescent Psychology, 38,* 899-908.

Mackinnon, S. P., Kehayes, I. L. L., Clark, R., Sherry, S. B., & Stewart, S. H. (2014). Testing the four-factor model of personality vulnerability to alcohol misuse: A three-wave, one-year longitudinal study. *Psychology of Addictive Behaviors, 28,* 1000.

Malhorta, N. (2008). Completion Time and Response Order Effects in Web Surveys. *Public Opinion Quarterly, 72*(5), 914-934.

Object Planet (2016). Opinio Survey Software [Computer Software]. Retrieved from http://www.objectplanet.com/opinio/

Podsakoff, P. M., MacKenzie, S. B., Lee, J. Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology, 88,* 879-903. doi:10.1037/0021-9010.88.5.879

Schwarz, N. (1999). Self-reports: How the questions shape the answers. *American Psychologist, 54,* 93-105.

Wilsnack, R. W., Wilsnack, S. C., Kristjanson, A. F., Vogeltanz- Holm, N. D., & Gmel, G. (2009). Gender and alcohol consumption: Patterns from the multinational GENACIS project. *Addiction, 104,* 1487-1500.

Woicik, P. A., Stewart, S. H., Pihl, R. O., & Conrod, P. J. (2009). The Substance Use Risk Profile Scale: A scale measuring traits linked to reinforcement-specific substance use profiles. *Addictive Behaviors, 34,* 1042-1055.

Table 1

Comparing ascending vs. descending conditions using Generalized Linear Models

Outcome	Distribution	Descending Mean (SD)	Ascending Mean (SD)	95% CI of Mean Difference	Wald Chi-Square	Cohen's d
AUDIT-C	Normal	4.24 (2.78)	3.88 (2.66)	[-0.20, 0.73]	$\chi^2(1) = 3.44$, $p = .06$	0.13
Social	Normal	9.60 (2.25)	9.39 (2.08)	[-0.11, 0.54]	$\chi^2(1) = 1.72$, $p = .19$	0.10
Enhancement	Normal	8.44 (2.10)	8.05 (2.13)	[0.08, 0.71]	$\chi^2(1) = 6.01$, $p = .01$	0.18
Conformity	Poisson	5.68 (2.50)	5.67 (2.59)	[-0.37, 0.39]	$\chi^2(1) = 0.001$ $p = .97$	0.00
Coping	Poisson	5.97 (2.68)	6.05 (2.70)	[-0.48, 0.32]	$\chi^2(1) = 0.16$, $p = .69$	-0.03
Hopelessness	Poisson	13.30 (3.62)	13.60 (3.85)	[-0.82, 0.22]	$\chi^2(1) = 1.27$, $p = .26$	-0.08
Anxiety Sensitivity	Normal	12.46 (3.02)	12.52 (3.09)	[-0.50, 0.36]	$\chi^2(1) = 0.96$ $p = .76$	-0.02
Sensation Seeking	Normal	16.04 (3.53)	15.19 (3.41)	[0.37, 1.34]	$\chi^2(1) = 11.90$, $p = .001$	0.24
Impulsivity	Normal	10.33 (2.73)	10.14 (2.58)	[-0.18, 0.56]	$\chi^2(1) = 1.04$, $p = .31$	0.07

Note. After listwise deletion within each analysis, Ns ranged from 691 to 791. Predictor variable is experimental condition (ascending vs. descending). Using nested model comparison, the best fitting distribution (i.e., Normal, Poisson, or Negative Binomial) for each outcome was chosen based on ΔBIC values (see Supplementary Table 1). Poisson models used a log link, but means are presented in the original metric so means may be more easily interpreted. Cells shaded grey are statistically significant with 95% confidence intervals that do not contain zero.

Table 2

Generalized linear models adding in covariates and interaction terms

Parameter Estimates									
Parameter	**AUDIT**	**Social**	**Enhance**	**Conform**	**Coping**	**Hopeless**	**Anx. Sens**	**Sens. Seek**	**Impulsivity**
Intercept	3.86	9.51	8.12	5.77	6.10	13.61	12.35	15.86	10.58
Condition	0.13	0.05	0.14	1.00	1.00	0.99	0.11	0.34	0.11
Age	-0.01	-0.07	-0.03	0.99	1.00	1.00	0.00	-0.03	0.02
Race	0.78	0.24	0.44	1.03	0.98	0.98	-0.18	0.04	-0.33
Sex	0.28	0.24	0.20	1.06	1.01	1.00	-0.38	0.45	0.30
Condition * Age	-0.02	0.02	-0.00	1.00	1.00	1.00	0.01	-0.02	-0.02
Condition * Race	0.03	0.21	0.11	1.03	1.00	0.99	-0.13	0.10	-0.17
Condition * Sex	-0.04	0.11	0.02	1.02	1.01	1.00	0.13	-0.06	-0.12
95% Confidence Intervals [lower, upper]									
Parameter	**AUDIT**	**Social**	**Enhance**	**Conform**	**Coping**	**Hopeless**	**Anx. Sens**	**Sens. Seek**	**Impulsivity**
Intercept	3.58, 4.13	9.28, 9.74	7.88, 8.35	5.50, 6.06	5.81, 6.41	13.25, 13.99	12.04, 12.66	15.53, 16.20	10.31, 10.86
Condition	-0.15, 0.40	-0.18, 0.29	-0.10, 0.38	0.95, 1.05	0.95, 1.05	0.96, 1.02	-0.19, 0.42	0.01, 0.68	-0.17, 0.38
Age	-0.05, 0.03	-0.10, -0.04	-0.07, 0.00	0.98, 0.99	1.00, 1.01	1.00, 1.01	-0.04, 0.05	-0.07, 0.01	-0.02, 0.07
Race	0.55, 1.00	0.04, 0.45	0.24, 0.65	0.99, 1.07	0.94, 1.02	0.96, .999	-0.42, 0.07	-0.24, 0.32	-0.55, -0.12
Sex	0.03, 0.53	0.06, 0.43	0.01, 0.39	1.02, 1.10	0.96, 1.05	0.98, 1.03	-0.66, -0.09	0.15, 0.76	0.05, 0.54
Condition * Age	-0.06, 0.02	-0.01, 0.05	-0.04, 0.03	1.00, 1.01	0.99, 1.00	1.00, 1.00	-0.04, 0.05	-0.06, 0.02	-0.07, 0.02
Condition * Race	-0.19, 0.25	0.001, 0.41	-0.09, 0.31	0.99, 1.07	0.96, 1.04	0.97, 1.02	-0.37 0.11	-0.18, 0.38	-0.39, 0.05
Condition * Sex	-0.29, 0.21	-0.08, 0.29	-0.17, 0.22	0.98, 1.06	0.96, 1.05	0.97, 1.02	-0.16, 0.41	-0.37, 0.25	-0.36, 0.13

Note. Cond = Experimental condition. B = Regression coefficient. After listwise deletion within each analysis, Ns ranged from 788 to 688. Values presented (Bs) are unstandardized parameter estimates for models. All categorical variables were effect-coded (descending = 1, ascending = -1; male = 1, female = -1; Caucasian = 1, non-Caucasian = -1). Because Poisson models (Conformity, Coping, and Hopelessness) used a log link, parameter estimates were exponentiated before reporting in this table; thus, parameter estimates for these 3 models can be interpreted as Incident Rate Ratios. Cells shaded grey are statistically significant with 95% confidence intervals that do not contain zero (for normal models) or do not contain 1.00 (for Poisson models).

Table 3
Spearman correlations and comparisons of correlation magnitude across conditions

Variable 1	Variable 2	r_s (descending)	r_s (ascending)	95% CI of the difference between r_s coefficients	
				Lower	Upper
AUDIT-C	Social	.43	.38	-.08	.18
AUDIT-C	Enhancement	.43	.43	-.13	.12
AUDIT-C	Conformity	.09	.02	-.09	.22
AUDIT-C	Coping	.30	.24	-.08	.20
AUDIT-C	Hopelessness	.11	.06	-.10	.20
AUDIT-C	Anxiety Sensitivity	-.02	-.02	-.15	.15
AUDIT-C	Sensation Seeking	.24	.18	-.08	.21
AUDIT-C	Impulsivity	.30	.17	-.02	.27
Social	Enhancement	.45	.54	-.20	.03
Social	Conformity	.16	.22	-.21	.09
Social	Coping	.29	.20	-.05	.23
Social	Hopelessness	.02	.05	-.19	.12
Social	Anxiety Sensitivity	.05	.06	-.17	.14
Social	Sensation Seeking	.15	.14	-.15	.15
Social	Impulsivity	.13	.14	-.17	.14
Enhancement	Conformity	.12	.09	-.12	.18
Enhancement	Coping	.29	.34	-.18	.09
Enhancement	Hopelessness	.01	.03	-.17	.13
Enhancement	Anxiety Sensitivity	.04	-.09	-.03	.28
Enhancement	Sensation Seeking	.29	.28	-.14	.15
Enhancement	Impulsivity	.16	.19	-.18	.12
Conformity	Coping	.26	.35	-.23	.05
Conformity	Hopelessness	.10	.21	-.26	.04
Conformity	Anxiety Sensitivity	.10	.24	-.29	.01
Conformity	Sensation Seeking	.01	.02	-.16	.14
Conformity	Impulsivity	.21	.19	-.13	.16
Coping	Hopelessness	.35	.33	-.12	.15
Coping	Anxiety Sensitivity	.15	.12	-.12	.18
Coping	Sensation Seeking	.15	.07	-.07	.24
Coping	Impulsivity	.29	.25	-.10	.19
Hopelessness	Anxiety Sensitivity	.14	.19	-.19	.10
Hopelessness	Sensation Seeking	-.09	-.16	-.08	.22
Hopelessness	Impulsivity	.30	.12	.04	.33
Anxiety Sensitivity	Sensation Seeking	-.25	-.19	-.21	.09
Anxiety Sensitivity	Impulsivity	.09	.08	-.15	.16
Sensation Seeking	Impulsivity	.22	.20	-.12	.17
Overall Correlation		.17	.16		

Note. n = 335 (descending) and n = 321 (ascending) after listwise deletion to exclude non-drinkers and missing data. Note also that the 95% CIs comparing the magnitude of correlation coefficients in this table are exploratory, and were not pre-registered
*p < .05, ** p < .01, *** p < .001.