
Article:
Lewin, J. orcid.org/0000-0003-3688-4920, Field, M. orcid.org/0000-0002-7790-5559 and Davies, E. orcid.org/0000-0003-3577-3276 (2024) Investigating the impact of 'dark nudges' on drinking intentions: a between groups, randomized and online experimental study. British Journal of Health Psychology, 29 (1). pp. 272-292. ISSN 1359-107X

https://doi.org/10.1111/bjhp.12698

Received: 28 December 2022 | Accepted: 26 September 2023

DOI: 10.1111/bjhp.12698

the british psychological society
promoting excellence in psychology

ARTICLE

Investigating the impact of 'dark nudges' on drinking intentions: A between groups, randomized and online experimental study

Joel Lewin[1,2] | **Matt Field**[3] | **Emma Davies**[1]

[1]The Centre for Psychological Research, Oxford Brookes University, Oxford, UK

[2]Oxford Institute of Clinical Psychology Training and Research, The University of Oxford, Oxford, UK

[3]Department of Psychology, University of Sheffield, Sheffield, UK

Correspondence
Joel Lewin, The University of Oxford, Warneford Hospital, Oxford OX3 7JX, UK.
Email: joel.lewin@hmc.ox.ac.uk

Abstract

Objectives: This study explored how 'dark nudges' (tactics used in alcohol industry-funded responsible drinking campaigns) affect drinking intentions, perceived source credibility and whether individual differences in perceptions of prototypical drinkers moderated these effects.

Design: Two 2×2 between-groups online experimental studies.

Methods: Study 1 (N = 164) presented three alcohol health messages per condition, comprising social norm (healthy/unhealthy ("dark nudge")) by frame (loss/gain). Study 2 (N = 229) presented one message per condition, comprising cancer causality (single cause/multiple causes (dark nudge)) by funding disclosure (disclosure/non-disclosure (dark nudge)). Outcomes were drinking intentions and perceived source credibility. Exploratory analyses considered prototype perceptions as a between-subjects moderator.

Results: No significant effects of message frame, social norm, fundi or multiple cancer causality arguments on drinking intentions were found. In Study 2, in the dark nudge multiple cancer causality conditions, perceived source credibility was high when funding was undisclosed, but significantly lower when it was disclosed. Exploratory analyses suggested effects were moderated by prototype similarity. In Study 1, higher perceived similarity to a heavy drinker and lower perceived similarity to a responsible drinker were associated with higher drinking intentions in the unhealthy norm/gain frame condition, but lower drinking intentions in the other conditions.

Conclusions: Framing, social norm, funding disclosure and multiple causality manipulations as tested in this study

did not exert a dark nudge effect on drinking intentions. However, the exploratory analyses suggest it could be hypothesised that the types of messages used in alcohol industry-funded responsible drinking campaigns may result in greater drinking intentions among those who identify more as heavy drinkers and less as responsible drinkers. Perceived prototype similarity may be an important moderator of the impact of alcohol health messages that warrants further research. Study 2 suggests disclosure of industry funding guides judgements of the credibility of sources of misleading messages about alcohol and cancer.

KEYWORDS

alcohol labels, credibility, dark nudges, message framing, prototypes, responsible drinking, social norms

BACKGROUND

Given that alcohol accounted for three million deaths worldwide in 2016 (World Health Organization [WHO], 2018), the WHO has targeted a 10% reduction in harmful alcohol use by 2025 (WHO, 2016). Nudges involve alterations to the choice architecture that harness systematic cognitive biases to guide behaviour in consumers' best interests without prohibiting options or changing financial incentives (Thaler & Sunstein, 2008). Nudge-type interventions have shown significant effects on alcohol purchases through product placement (Nakamura et al., 2014) and warning labels (Zhao et al., 2020). Researchers have focused on optimizing alcohol warning labels and health messages through manipulations of social norms (Park et al., 2020) and framing (Churchill et al., 2016).

However, nudges may also be harnessed to encourage choices that benefit the industry and harm consumers (Newall, 2019). Newall (2013), who first described these strategies as 'dark nudges', identified selective use of fonts in industry-led responsible gambling initiatives, which subvert the content of the warning messages presented (Newall et al., 2022). Experimental findings suggest that alcohol branding amplifies cognitive bias in alcohol purchasing decisions (Pennington et al., 2022).

Petticrew et al. (2020) suggested that dark nudges are prevalent in responsible drinking campaigns conducted by alcohol industry (AI) funded corporate social responsibility (CSR) groups, or Social Aspects/Public Relations Organizations (SAPROs; Babor, 2009). These groups state objectives such as reducing harmful drinking and claim independence from the industry that funds them (Drinkaware, n.d; Drinkwise, n.d.a). However, studies have suggested that their activities further AI interests, often at the expense of public health (Babor et al., 2018; McCambridge et al., 2014; Petticrew et al., 2016). Petticrew et al.'s (2020) review paper documents strategies such as framing alcohol consumption by its benefits rather than its costs, as seen in a Canadian SAPRO article listing "8 benefits of moderate drinking" (Éduc'alcool, 2017). The article suggests, "sticking to the recommended drinking guidelines will help you keep your weight under control!", a frame that obscures the fact that the opposite is true compared with not drinking at all. Petticrew et al.'s (2020) argument is supported by experimental findings that participants reported significantly lower motivation to reduce alcohol consumption (Brennan et al., 2020) and drank significantly more in both a laboratory and bar context (Moss et al., 2015) after viewing SAPRO harm reduction advertisements compared with public health agency advertisements. A qualitative (Pettigrew, Biagioni, et al., 2016) and a quantitative study (Smith et al., 2006) both found that participants interpreted industry-funded responsible drinking advertisements as encouraging drinking and making it look fun. Smith et al. (2006) concluded that advertisements use strategic ambiguity to present ostensibly pro-health messages while actually benefiting industry sales and public relations.

Statement of Contribution

What is already known on this subject?

- Alcohol industry-funded responsible drinking campaigns fail to change drinking behaviours.
- Strategic use of social norms and framing in alcohol health messages influences drinking intentions and behaviours.
- Exposure to industry-sponsored messages increases uncertainty or false certainty about the links between alcohol and breast cancer.
- Knowledge of the source of funding for industry-sponsored health messages influences credibility judgements.

What does this study add?

- There was no overall significant dark nudge effect of framing, social norm or alternative causation argument on drinking intentions.
- Perceived credibility of the source of alcohol health messages obfuscating the links between alcohol and cancer is lower when industry funding is disclosed.
- Exploratory analyses suggested people who perceive themselves as more similar to a heavy drinker and less similar to a responsible drinker may be more vulnerable to dark nudge alcohol health messages.
- Exploratory analyses suggested prototype perceptions may moderate the impact of dark nudge messages on drinking intentions and should be explored further.

Social norms and framing

Some experiments suggest gain-framed messages, highlighting the benefits of behaviour change, have a greater impact on drinking intentions (Park et al., 2020), especially among heavier drinkers (Kingsbury et al., 2015), whereas others suggest loss-framed messages, highlighting the costs of not changing a behaviour, are associated with greater motivation to drink less (Blackwell et al., 2018).

The social norm effect is hypothesised to work by nudging people to adjust their behaviour to conform to the presented norm (Thaler & Sunstein, 2008). Park et al. (2020) found that messages conveying healthy norms ('The majority of college students drink an average of two or fewer drinks a week') had the greatest effect on intentions to drink 'responsibly' (two or fewer drinks a week), which was amplified by a gain frame. Conversely, messages combining an unhealthy norm ('The majority of college students drink an average of five or more drinks a week.') with a loss frame were associated with the lowest responsible drinking intentions. This suggests a focus on unhealthy norms could serve as a dark nudge undermining the efficacy of responsible drinking messages, an effect that may be amplified by interactions with framing effects.

AI-funded CSR campaigns have framed survey data to focus on people who increased alcohol consumption during the COVID-19 pandemic, even when the same data shows more people reduced consumption (Drinkaware, 2020). For example:

> Our new research out today reveals substantial patterns in the nation's drinking over the lockdown. One in 10 drinkers – equivalent to 4.6 million in the UK – drank more than normal throughout lockdown (March to August).
>
> (Drinkaware, 2020)

This may encourage compliance with the unhealthy behaviour and diminish the salience of the healthy behaviour. The unhealthy norm frame has sometimes been combined with a gain frame highlighting the

benefits of drinking, for example, "50% report 'coping' as a reason for drinking alcohol" (Drinkaware IRL, 2020). As such, Study 1 sought to explore the effects of framing and social norms on drinking intentions in the types of messages used in AI SAPRO campaigns.

Cancer and funding disclosure

It has been suggested (Petticrew et al., 2018) that AI-funded campaigns misrepresent and omit evidence on the links between alcohol and cancer (particularly colorectal and breast cancer), presenting confounding misinformation that downplays risks and casts doubt on causality with "alternative causation" arguments (Maani et al., 2022) that emphasize risk factors other than alcohol. Drinkwise's website, for example, states:

> Not all heavy drinkers get cancer as multiple risk factors are involved in the development of cancers including genetics and family history of cancer, age, environmental factors and behavioural variables, as well as social determinants of health.
>
> (Drinkwise, n.d.b)

Maani et al. (2022) found that exposure to industry-sponsored alternative causation messages significantly increased uncertainty or false certainty about the links between alcohol and breast cancer compared with exposure to non-industry-sponsored messages.

This is pertinent given that unambiguous cancer warning messages are associated with a greater likelihood to consider drinking less (Winstock et al., 2020). A quasi-experiment introduced warning labels, including messages highlighting links between alcohol and breast and colorectal cancer, to alcoholic beverages in one region of Canada (Zhao et al., 2020). The study was paused after one month following AI complaints about the cancer messages and was only permitted to resume without the cancer labels (Stockwell et al., 2020). Nonetheless, there was a significant increase in participants' awareness of links between alcohol and cancer (Hobin et al., 2020) and a population-level decline in alcohol consumption in the region (Zhao et al., 2020). This study therefore aimed to examine the impact of alcohol industry-sponsored alternative causation messages.

Source credibility, comprising perceived expertise and trustworthiness, influences a message's impact on explicit and implicit evaluations of commercial products (Smith et al., 2013) and on alcohol consumption (Harris et al., 2009). However, the discounting principle suggests a source's credibility can be undermined by disclosure of conflicts of interest such as payment (Sparkman Jr., 1982). This is potentially relevant insofar as SAPRO spokespeople, such as the members of Drinkaware's Medical Advisor Panel who frequently appear in campaign materials, receive an undisclosed annual retainer from Drinkaware (Sim et al., 2018). Brennan et al. (2017) found that the 84.1% of respondents who incorrectly believed Australian SAPRO Drinkwise received government funding had a more favourable perception of Drinkwise's trustworthiness, credibility and respectability than the 37% who were aware it received industry funding.

Prototype perceptions

Health interventions affect different people differently, yet moderators are poorly understood in alcohol labelling interventions (Hassan & Shiu, 2018). The social identity component of the Prototype Willingness Model (PWM; Davies & Todd, 2021) suggests people hold distinctive images of the type of person who engages in certain behaviours, and that one is more likely to engage in that risk behaviour if that prototypical image is evaluated favourably and perceived as more similar to oneself. Perceptions of prototypical heavy drinkers or non-drinkers are significant predictors of drinking behaviour (Davies, 2019; Davies & Todd, 2021), yet little research has explored their role as moderators of

message impact. Davies et al. (2022) found that the similarity and favourability of a responsible drinker prototype moderated the effect of different message frames on future drinking intentions. The current study therefore examined whether prototype perceptions moderate the impact of dark nudges on drinking intentions.

To summarize, AI-funded responsible drinking campaigns are less effective than public health campaigns (Brennan et al., 2020), possibly due to dark nudges that undermine the effect of industry-funded campaigns (Petticrew et al., 2020). However, there is little experimental research testing specific dark nudge mechanisms. Examination of descriptive reviews of AI-funded materials and previous research on alcohol health message efficacy led to the following pre-registered (https://osf.io/yun37/) hypotheses.

Study 1 primary hypotheses

H1. Healthy norm messages stating, "1 in 4 adults are drinking less since COVID-19" will be associated with lower overall drinking intentions than dark nudge unhealthy norm messages stating "1 in 4 adults are drinking more since COVID-19".

H2. Gain frame messages will affect overall drinking intentions differently to loss frames. No direction is specified as previous findings are mixed.

H3. There will be an interaction between social norm and frame. No direction is specified as previous research is limited.

Study 2 primary hypotheses

H4. Messages with no disclosure of industry funding (dark nudge) will be associated with higher drinking intentions than messages with disclosure.

H5. Single causality messages will affect drinking intentions differently to dark nudge multiple causality messages. No direction was specified in the protocol, but based on the research published by Maani et al. (2022), it should be hypothesised that single causality messages will be associated with lower drinking intentions than multiple causality messages.

H6. There will be an interaction effect between cancer causality and disclosure of funding. More specifically, multiple causality will be associated with higher drinking intentions when combined with non-disclosure (dark nudge), but with lower drinking intentions when combined with disclosure.

Study 2 secondary hypothesis

H7. Perceived credibility will be higher and overall drinking intentions lower when funding disclosure is combined with a single causality message, but the opposite effects will be seen when funding disclosure is combined with a multiple causality message (dark nudge).

Study 1 and 2 exploratory hypothesis

H8. Perceived prototype familiarity and similarity will moderate the effect of messages on overall drinking intentions.

METHODS

Design and procedure

Both Studies 1 and 2 used a 2 × 2 between-groups experimental design conducted online with Qualtrics survey software.

Study 1 comprised social norm (healthy/unhealthy) by frame (loss/gain). Study 2 comprised cancer causality (multiple causes/single cause) by funding disclosure (disclosure/non-disclosure).

Participants read a participant information sheet and confirmed consent. The information sheet did not mention dark nudges (see Supporting Information 3).

Participants submitted demographic information, alcohol consumption and prototype perceptions, before being randomized. Participants were instructed to read the messages carefully and think about the information.

In Study 1, participants viewed three messages, each displayed separately for 15 s, before completing drinking intention measures (see Figure 1). In Study 2, participants viewed one message for 30 s, then completed credibility measures on the same page and drinking intention measures on the next page (see Figure 2). The final debrief page gave participants the option to withdraw consent.

Ethical approval was granted by the Psychology Research Ethics Committee, Oxford Brookes University (reference number 2021/321). Hypotheses, methods and analyses were pre-registered on the Open Science Framework (https://osf.io/yun37/).

Participants

Participants aged 18 and over who drink alcohol at least once a fortnight and currently reside in the UK were recruited opportunistically through social media (see Supporting Information 1 for recruitment messages).

In total, 128 participants per study were required to provide 80% power for a medium effect size ($f = .25$) and an alpha level of .05, as calculated using Gpower. An extra 20% (26 participants per study) were added to account for possible incomplete responses. Therefore, each study aimed to recruit 154 participants.

In Study 1, 59 responses were excluded because they did not complete the survey and four withdrew consent on the debrief page, leaving 164 complete responses (54.9% women; mean age = 40.06, $SD = 13.35$, range = 21–75 years). In Study 2, 10 were excluded as incomplete and two withdrew consent, leaving 229 participants (61.6% women; mean age = 41.83, $SD = 13.81$, range = 19–75 years). (See Tables S1 and S2 for full demographic data.)

In both studies, there were no significant differences between groups in terms of age, AUDIT-C scores or prototype perceptions, indicating randomisation was successful.

Stimuli

Message stimuli were created to imitate formats commonly used by SAPROs such as Drinkaware on social media (Drinkaware, 2020). See details in Supporting Information 2.

FIGURE 1 Study 1 unhealthy norm/gain frame messages.

In Study 1, the social norm message was the same for all three messages in each condition, with only one word changed between the healthy/unhealthy conditions: "1 in 4 adults are drinking less/more since COVID-19".

The second part of the message stated a "top reason" why people engaged in the normative behaviour, in terms of either the positives gained (gain frame), or negatives avoided (loss frame). Three messages per condition were used to avoid the problem of stimulus sampling (Wells & Windschitl, 1999). Each condition presented a message about mood and stress, sleep and loneliness, in that order (see Figure S1 for Study 1 stimuli).

In Study 2, the cancer causality variable had two conditions. The multiple causes condition described alcohol as one of many risk factors for cancer and emphasized the lack of scientific consensus. The single-cause condition focused only on alcohol as a cause of cancer.

The funding disclosure variable was manipulated by either including (disclosure) or excluding (non-disclosure) a funding disclosure statement.

FIGURE 2 Study 2 messages for multiple causality/ disclosure condition, and single causality/non-disclosure condition.

A fictitious SAPRO name, 'Drinkalert' was created to avoid any confounding effects of pre-existing perceptions (Atkin et al., 2008). Messages are included in Figure S2.

Measures

Main outcome measures

Intentions

Behavioural intentions were measured in line with previous research on alcohol health messages and drinking intentions (Pettigrew, Jongenelis, et al., 2016).

Drinking intentions. Participants were asked: To what extent do you intend to do the following in the next two weeks:

- Have an alcoholic drink? (from 1 = 'definitely do not intend to do this' to 7 = 'definitely intend to do this').
- Get drunk?

These two scores were averaged to produce a drinking intention score.

Reduction intentions. Participants were asked:

- To what extent do you believe you should reduce the amount of alcohol you consume? (from 1 = not at all to 7 = to a great extent)
- To what extent do you think you will actually reduce the amount of alcohol you consume?

These two scores were averaged to produce a reduction intentions score.

Overall drinking intentions. A composite overall drinking intentions score was produced by reversing reduction intentions and averaging it with drinking intentions. This composite score is the primary outcome variable used for confirmatory analysis. In Study 1 $\alpha = .278$ and in Study 2 $\alpha = .272$.

Credibility. In Study 2, perceived credibility was measured using 7-point semantic differential scale items in line with previous research on perceptions of AI-funded CSR groups (Brennan et al., 2017).

Participants were asked: Please rate Drinkalert for how [credible]/[trustworthy]/[expert] you think they are. [1 = not credible – 7 = credible]/[1 = untrustworthy – 7 = trustworthy]/[1 = not expert – 7 = expert].

A composite credibility score was calculated by averaging the scores for credibility, expertise and trustworthiness (3 items, $a = .931$). This was used as the outcome variable and is henceforth referred to as perceived credibility.

Theory-based measures

Prototypes

In line with previous research (Davies, 2019) respondents were asked to think about someone the same age as them who is a typical heavy drinker and a typical responsible drinker. For each prototype participants rated favourability (1 = extremely negative; 7 = extremely positive) and perceived similarity to that imagined person (1 = not at all; 7 = very; see Supporting Information 3 for full instructions).

Demographics

The abbreviated three-question Alcohol Use Disorders Identification Test Consumption survey (AUDIT-C) was used to measure alcohol consumption (Reinert & Allen, 2007). Participants also completed information about their gender, age, ethnicity, education and occupation.

Analysis

Two-way ANOVAs were used to test for main effects and interactions in terms of intention variables in both studies and composite credibility in Study 2. Assumptions for two-way ANOVA analyses were met. Post-hoc pairwise comparisons were conducted using Tukey's Honestly Significant Difference Test to explore significant effects. *p*-values were adjusted for multiple comparisons using Tukey's adjustment.

The PROCESS macro (Hayes, 2018) in SPSS was used to examine whether prototype perceptions moderated the effect of message condition on overall drinking intentions. Message condition was entered as the independent multicategorical predictor variable (X) using indicator coding, which creates three variables with one baseline group compared with each other group. The baseline group was unhealthy norm/gain frame in Study 1 and single causality/non-disclosure in Study 2. Overall drinking intentions were entered as the outcome variable (Y) and the prototype perception variable was centred and entered as the moderator variable (W). Separate moderation analyses were conducted for each prototype perception variable. Interactions of $p < .1$ were explored visually and probed by comparing the differences between simple slopes by the condition at mean, high ($M + 1SD$) and low ($M - 1SD$) levels of the prototype perception variable.

TABLE 1 Descriptive statistics and ANOVA test statistics for Study 1 and Study 2.

	n	Statistic	Overall drinking intentions	Reduction intentions	Drinking intentions	Composite credibility
Study 1						
Unhealthy gain	41	*M* (*SD*)	5.09 (.84)	2.82 (1.42)	4.99 (1.39)	
Unhealthy loss	39	*M* (*SD*)	4.79 (1.17)	3.4 (1.72)	4.99 (1.26)	
Healthy gain	42	*M* (*SD*)	4.7 (.96)	3.3 (1.52)	4.7 (1.22)	
Healthy loss	42	*M* (*SD*)	4.65 (1)	3.39 (1.73)	4.69 (1.69)	
Social norm		F, p, η_p^2	$F = 2.887, p = .091, \eta_p^2 = .018$	$F = .904, p = .343, \eta_p^2 = .006$	$F = 1.76, p = .187, \eta_p^2 = .011$	
Frame		F, p, η_p^2	$F = 1.221, p = .271, \eta_p^2 = .008$	$F = 1.821, p = .179, \eta_p^2 = .011$	$F = .001, p = .977, \eta_p^2 = 0$	
Social Norm × Frame		F, p, η_p^2	$F = .579, p = .448, \eta_p^2 = .004$	$F = .939, p = .334, \eta_p^2 = .006$	$F = .001, p = .98, \eta_p^2 = 0$	
Study 2						
Multiple causality/disclosure	56	*M* (*SD*)	4.42 (.96)	3.5 (1.56)	4.34 (1.5)	2.88 (1.36)
Multiple causality/ non-disclosure	58	*M* (*SD*)	4.51 (.92)	3.58 (1.71)	4.6 (1.33)	4.47 (1.68)
Single causality/disclosure	56	*M* (*SD*)	4.62 (1.28)	3.34 (1.72)	4.57 (1.57)	3.96 (1.63)
Single causality/non-disclosure	59	*M* (*SD*)	4.48 (.88)	3.58 (1.46)	4.55 (1.57)	4.15 (1.41)
Cancer causality		F, p, η_p^2	$F = .382, p = .537, \eta_p^2 = .002$	$F = .129, p = .72, \eta_p^2 = .001$	$F = .207, p = .65, \eta_p^2 = .001$	$F = 3.532, p = .061, \eta_p^2 = .015$
Disclosure		F, p, η_p^2	$F = .022, p = .883, \eta_p^2 = 0$	$F = .572, p = .45, \eta_p^2 = .003$	$F = .38, p = .538, \eta_p^2 = .002$	**$F = 19.29, p < .001, \eta_p^2 = .079$**
Multiple causality × Disclosure		F, p, η_p^2	$F = .705, p = .402, \eta_p^2 = .003$	$F = .155, p = .695, \eta_p^2 = .001$	$F = .52, p = .472, \eta_p^2 = .002$	**$F = 11.957, p = .001, \eta_p^2 = .05$**

Bold values indicate $p < .05$.

RESULTS

Study 1

Descriptive statistics

Overall drinking intentions were highest in the unhealthy norm/gain frame condition ($M = 5.09$, $SD = .84$) and lowest in the healthy norm/loss frame condition ($M = 4.65$, $SD = 1$; see Table 1).

Primary hypotheses

The main effect of social norm on drinking intentions was not significant, $F(1, 160) = 2.89$, $p = .091$, $\eta_p^2 = .018$ (see Table 1 and Figure 1). Neither was there a significant main effect of frame, $F(1, 160) = 1.22$, $p = .271$, $\eta_p^2 = .008$, nor a significant interaction between social norm and frame, $F(1, 160) = .58$, $p = .448$, $\eta_p^2 = .004$. As such, H1, H2 and H3 were not supported.

Exploratory analyses

Prototype perception moderation analyses

Neither heavy drinker favourability nor responsible drinker favourability significantly moderated the effect of message conditions on drinking intentions.

However, heavy drinker and responsible drinker similarity both significantly moderated the effect of message condition on drinking intentions (see Table 2). Figure 3 shows that at low levels of heavy drinker similarity, overall drinking intentions were similar across conditions, with no significant differences in simple slopes. However, higher levels of perceived heavy drinker similarity and lower perceived responsible drinker similarity were associated with higher drinking intentions in the unhealthy norm/gain frame condition, but lower drinking intentions in the other conditions. At high levels of heavy drinker similarity (4.97), simple slopes comparison indicated the effect of the unhealthy norm/gain frame was significantly different to the healthy norm/loss frame ($b = -.86$, $t = -2.76$, $p = .006$) and healthy norm/gain frame ($b = -.70$, $t = -2.21$, $p = .029$).

The inverse effect was seen for responsible drinker similarity (see Figure 4). At lower levels of responsible drinker similarity (2.95), the effect of the unhealthy norm/gain frame was significantly different to the healthy norm/loss frame ($b = -.75$, $t = -2.32$, $p = .022$) and unhealthy norm/loss frame ($b = -.64$, $t = -2.09$, $p = .038$).

Study 2

Descriptive statistics

Overall drinking intentions were similar across conditions, but highest in the single causality/disclosure condition ($M = 4.62$. $SD = 1.28$) and lowest in the multiple causality/disclosure condition ($M = 4.42$, $SD = .96$; see Table 1).

Perceived credibility was lowest in the multiple causality/disclosure condition ($M = 2.88$, $SD = 1.36$) and highest in the multiple causality/non-disclosure condition ($M = 4.47$, $SD = 1.68$; see Table 1).

TABLE 2 Results of Study 1 exploratory moderation analyses exploring whether perceived similarity to a heavy drinker or responsible drinker moderated the impact of the condition on drinking intentions with unhealthy norm/gain frame messages as the reference group.

	Coefficient	*SE*	*t*	*p*	Lower CI	Upper CI
Heavy drinker similarity						
Constant	5.12	.16	32.55	<.001	4.81	5.43
Unhealthy norm/loss frame	−.32	.22	−1.43	.154	−.76	.12
Healthy norm/gain frame	−.41	.22	−1.86	.064	−.84	.02
Healthy norm/loss frame	−.45	.22	−2.05	.042	−.89	−.02
Heavy drinker similarity	.14	.1	1.36	.176	−.06	.33
Heavy drinker similarity × Unhealthy norm/loss frame	−.16	.16	−1.04	.301	−.47	.15
Heavy drinker similarity × Healthy norm/gain frame	−.18	.13	−1.34	.183	−.44	.09
Heavy drinker similarity × Healthy norm/loss frame	−.25	.13	−1.94	.054	−.51	.00
Responsible drinker similarity						
Constant	5.10	.16	32.83	<.001	4.79	5.40
Unhealthy norm/loss frame	−.22	.22	−.98	.331	−.66	.22
Healthy norm/gain frame	−.40	.22	−1.86	.065	−.84	.03
Healthy norm/loss frame	−.45	.22	−2.08	.039	−.88	−.02
Responsible drinker similarity	−.06	.10	−.56	.576	−.26	.14
Responsible drinker similarity × Unhealthy norm/loss frame	.27	.14	1.92	.057	−.01	.55
Responsible drinker similarity × Healthy norm/gain frame	.14	.14	.96	.339	−.15	.42
Responsible drinker similarity × Healthy norm/loss frame	.19	.14	1.33	.185	−.09	.48

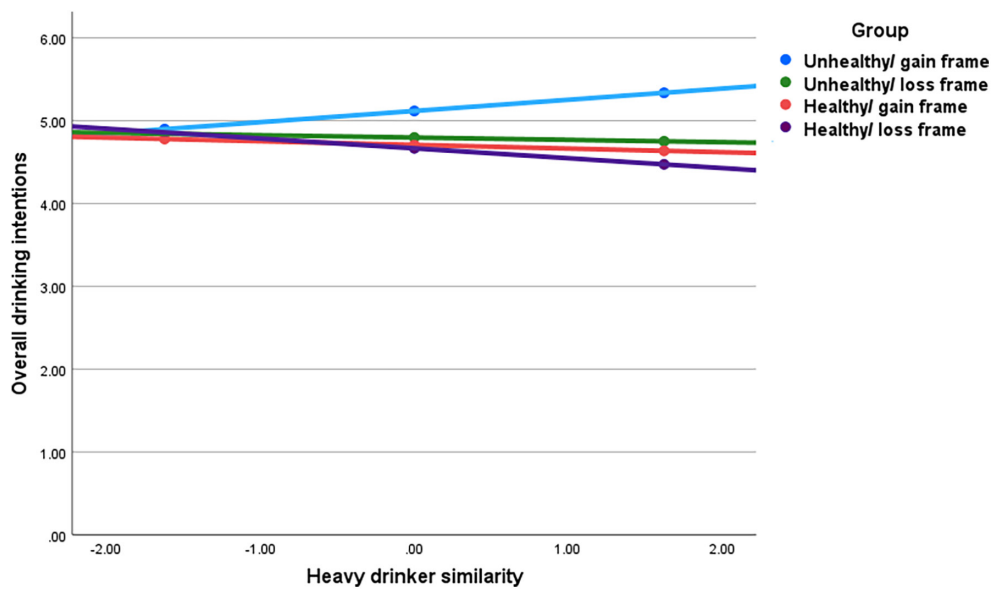

FIGURE 3 Overall drinking intentions by condition with heavy drinker similarity (mean-centred) as moderator at three levels- mean and 1*SD* above and below.

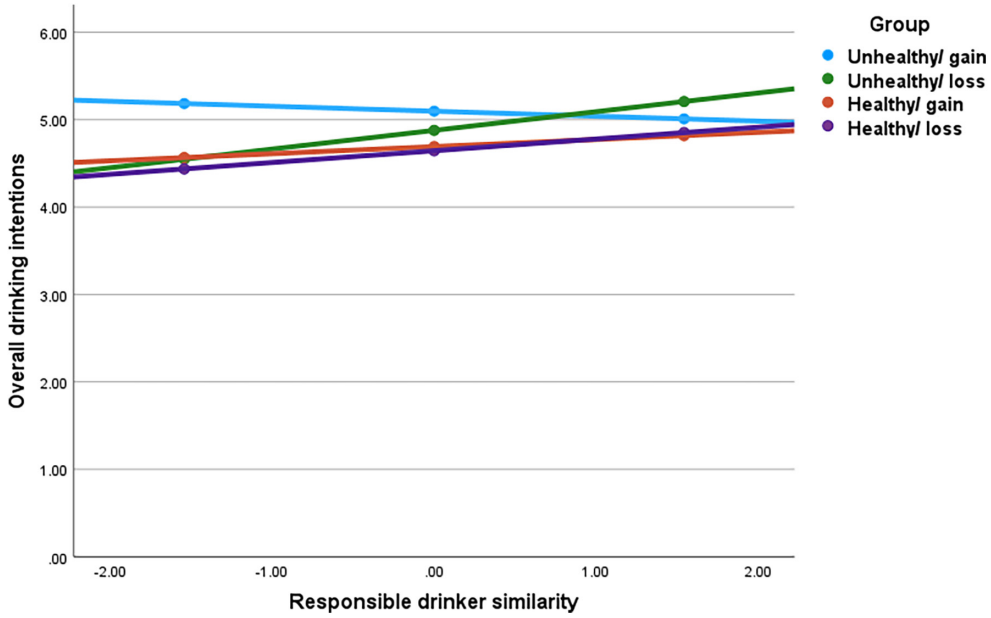

FIGURE 4 Overall drinking intentions by condition with responsible drinker similarity (mean-centred) as moderator at three levels- mean and 1*SD* above and below.

Primary hypotheses

There was no significant main effect of cancer causality on overall drinking intentions, $F(1, 225) = .38$, $p = .537$, $\eta_p^2 = .002$, or disclosure, $F(1, 225) = .02$, $p = .883$, $\eta_p^2 = .000$ and no significant interaction effect, $F(1, 225) = .71$, $p = .402$, $\eta_p^2 = .003$ (see Table 1). As such, H4, H5 and H6 were not supported.

Secondary hypothesis

Perceived credibility

There was a significant small-to-medium size interaction effect between causality and disclosure in terms of perceived credibility, $F(1, 225) = 11.96$, $p = .001$, $\eta_p^2 = .050$ and a significant medium size main effect of disclosure, $F(1, 225) = 19.29$, $p < .001$, $\eta_p^2 = .079$, but no significant main effect of causality, $F(1, 225) = 3.53$, $p = .061$, $\eta_p^2 = .015$ (see Table 1 and Figure S3).

Post-hoc tests showed perceived credibility was significantly lower in the multiple causality/disclosure condition ($M = 2.88$, $SD = 1.36$) than in the multiple causality/non-disclosure condition ($M = 4.47$, $SD = 1.68$, $p < .001$), the single causality/disclosure condition ($M = 3.96$, $SD = 1.63$, $p = .001$) and the single causality/non-disclosure condition ($M = 4.15$, $SD = 1.41$, $p < .001$). This partially supports H8 (see Table S3).

Exploratory analyses

Prototype perceptions moderation analyses

Neither heavy drinker nor responsible drinker favourability moderated the effect of conditions on overall drinking intentions. However, perceived similarity to both a heavy and a responsible drinker moderated the relationship between message condition and overall drinking intentions (see Table 3).

TABLE 3 Results of Study 2 exploratory moderation analyses exploring whether perceived similarity to a heavy drinker or responsible drinker moderated the impact of the condition on drinking intentions with single causality/non-disclosure messages as the reference group.

	Coefficient	*SE*	*t*	*p*	Lower CI	Upper CI
Heavy drinker similarity						
Constant	4.46	.13	33.69	<.001	4.20	4.72
Multiple causality/disclosure	−.04	.19	−.19	.848	−.41	.34
Multiple causality/non-disclosure	.05	.19	.27	.790	−.32	.42
Single causality/disclosure	.11	.19	.60	.549	−.26	.49
Heavy drinker similarity	.10	.08	1.35	.179	−.05	.25
Heavy drinker similarity × Multiple causality/disclosure	−.15	.11	−1.35	.178	−.36	.07
Heavy drinker similarity × Multiple causality/non-disclosure	−.23	.11	−2.14	**.033**	−.44	−.02
Heavy drinker similarity × Single causality/disclosure	−.31	.11	−2.69	**.008**	−.53	−.08
Responsible drinker similarity						
Constant	4.46	.13	33.42	<.001	4.20	4.72
Multiple causality/disclosure	−.04	.19	−.21	.833	−.42	.33
Multiple causality/non-disclosure	.05	.19	.28	.779	−.32	.42
Single causality/disclosure	.12	.19	.64	.526	−.26	.50
Responsible drinker similarity	−.10	.09	−1.16	.247	−.27	.07
Responsible drinker similarity × Multiple causality/disclosure	.09	.13	.68	.495	−.16	.34
Responsible drinker similarity × Multiple causality/non-disclosure	.22	.12	1.83	.069	−.02	.45
Responsible drinker similarity × Single causality/disclosure	.25	.12	2.07	**.040**	.01	.49

Bold values indicate $p < .05$.

Figure 5 shows that greater perceived heavy drinker similarity was associated with lower drinking intentions relative to lower perceived heavy drinker similarity in all conditions except for single causality/non-disclosure, in which the opposite was seen. Simple slopes analysis indicated that at low levels of heavy drinker similarity, there was a significant difference in the effect of single causality/non-disclosure compared with single causality/disclosure ($b = -.64$, $t = 2.35$, $p = .02$), with drinking intentions greater in the single causality/non-disclosure group at higher levels of heavy drinker similarity, but lower in the single causality/disclosure condition.

Symmetrical moderation effects were seen for responsible drinker similarity (see Figure 6). Simple slopes analysis indicated that at high levels of responsible drinker similarity, the difference in the effect of single causality/non-disclosure compared with single causality/disclosure approached statistical significance ($b = .51$, $t = 1.91$, $p = .058$).

DISCUSSION

Study 1

There were no significant effects of social norm or frame on drinking intentions and no significant interaction. The primary hypotheses were therefore not supported.

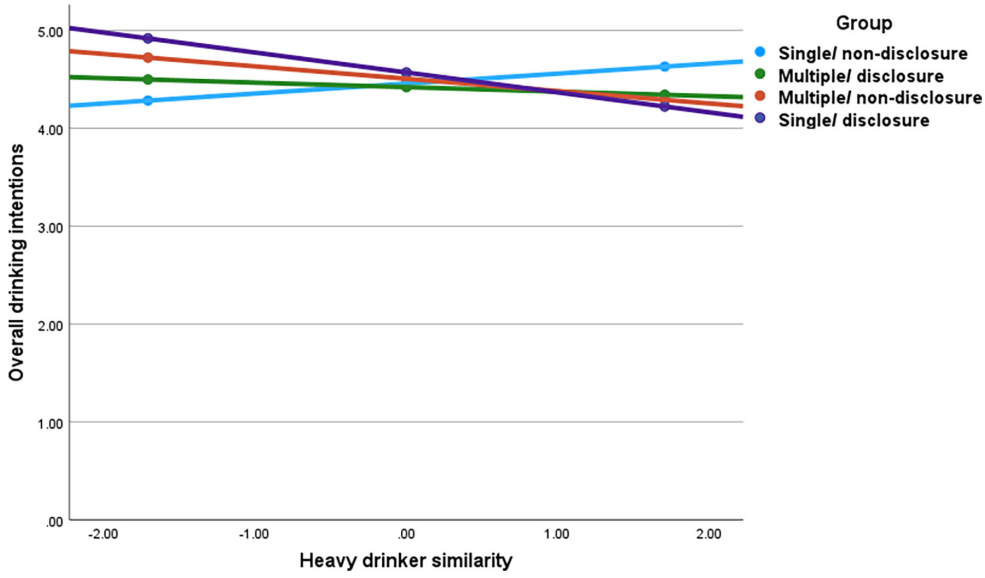

FIGURE 5 Overall drinking intentions by condition with heavy drinker similarity (mean-centred) as moderator at three levels- mean and 1*SD* above and below.

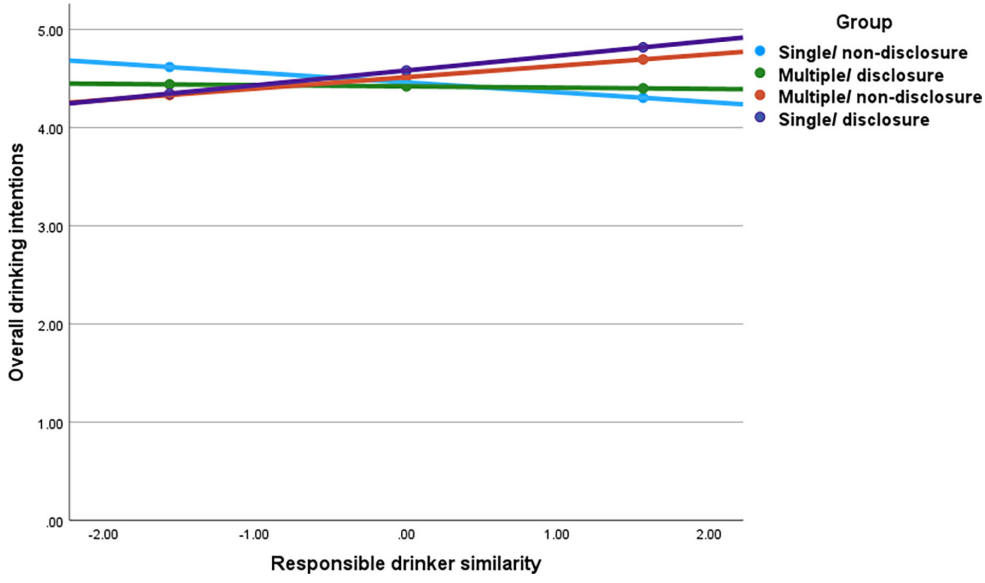

FIGURE 6 Overall drinking intentions by condition with responsible drinker similarity (mean-centred) as moderator at three levels- mean and 1*SD* above and below.

This could suggest that social norm and framing manipulations do not serve as dark nudges in AI-funded campaigns. Concerns about small effect sizes in social norm interventions have been raised (Foxcroft et al., 2015).

However, it may reflect the fact this study explored the suboptimal (Brennan et al., 2020) messages featured in industry-funded campaigns, in contrast to typical alcohol health message studies that seek to identify messages optimal for intentions and behaviour change (Blackwell et al., 2021). The messages in

this study required themes that could be framed as both costs and benefits of drinking more and drinking less, and therefore could not include themes that are most effective in reducing drinking, such as specific diseases (Dimova & Mitchell, 2022). As such, the healthy norm messages were essentially hamstrung. Matching the message content allowed direct comparison of mechanisms without confounding effects of different content, but does not offer insights into the differential effects of AI-funded campaign messages compared with optimally designed messages, which would also differ in content. Studies comparing actual SAPRO messages with public health messages have found lower motivation to reduce alcohol consumption (Brennan et al., 2020) and greater alcohol consumption (Moss et al., 2015) after exposure to SAPRO messages compared with public health messages. Moss et al. (2015) exposed participants to actual SAPRO posters in a simulated bar environment, a naturalistic setting that contrasts with the artificial experimental context of the present study, another possible explanation for the present null findings.

Study 2

There were no significant effects of message disclosure or cancer causality on drinking intentions and no interaction. As such, the primary hypotheses were not supported.

However, there was a significant main effect of disclosure and a significant interaction between disclosure and cancer causality in terms of perceived source credibility. In the clear, factual, single causality condition, there was no difference in perceived credibility when funding was disclosed or undisclosed. However, in the "dark nudge" multiple causality condition, when AI funding was disclosed, perceived credibility was significantly lower than in all other conditions. This is consistent with the discounting principle, by which the perceived trustworthiness of a spokesperson is diminished by disclosure of payment (Sparkman Jr., 1982). It is also consistent with findings that alcohol company sponsorship of an anti-drink driving message leads people to infer ulterior motives (Szykman et al., 2004), and that for ambiguous smoking cessation messages, markers of industry sponsorship result in lower perceived credibility (Byrne et al., 2012).

Therefore, this work builds on other recent work exploring the effect of the transparency of AI funding and the credibility of alcohol-related marketing information (Brennan et al., 2017; Petticrew et al., 2020).

Prototype perceptions

Although both studies found no overall significant effects of the messages on drinking intentions, the exploratory moderation analyses found effects conditional on prototype similarity.

In Study 1, greater perceived similarity to a heavy drinker prototype and lower perceived similarity to a responsible drinker prototype were associated with lower drinking intentions in the healthy norm conditions, but higher drinking intentions in the unhealthy norm/gain frame condition, suggesting these participants were more vulnerable to this dark nudge. In theory, these participants should benefit most from a responsible drinking campaign, given perceived personal relevance predicts the impact of alcohol health messages (Pettigrew, Jongenelis, et al., 2016). Indeed, these participants benefitted most from healthy norm messages, yet were most adversely impacted by AI campaign-style messages presenting an unhealthy norm and the benefits of drinking. Given findings of enhanced recall for feedback that confirms one's self-conception (Swann & Read, 1981), these dark nudge messages possibly reinforced their identity as heavy drinkers and justified their unhealthy behaviours.

In Study 2, exploratory analyses suggested that, in the clear, factual single cancer causality conditions, greater perceived heavy drinker similarity and lower responsible drinker similarity were associated with lower drinking intentions when funding was disclosed, but higher drinking intentions when funding was undisclosed. This may be explained by findings that low-prestige sources can be highly persuasive

when arguing against their self-interest (Walster et al., 1966), and that sources can be more persuasive when presenting arguments contrary to those which participants would expect them to make (Eagly & Chaiken, 1975). It could be hypothesised from these exploratory findings that among participants who perceive themselves as more similar to a heavy drinker prototype and less similar to a responsible drinker prototype, messages clearly describing the links between alcohol and cancer would be associated with lower drinking intentions when industry funding is disclosed, while exposure to messages focusing on unhealthy norms and the benefits of drinking alcohol would be associated with higher drinking intentions, echoing Petticrew et al.'s (2020) suggestions, but specifically for this group. Future research should test these hypotheses.

The present study builds on previous work demonstrating the important moderating role played by prototype willingness (Davies et al., 2022).

Strengths and limitations

This study only measured perceived credibility and intentions as outcome variables. Behavioural intentions are a useful but imperfect predictor of health behaviours (Webb & Sheeran, 2006). The composite score in this study had poor internal reliability. In addition, the first reduction intentions item—"To what extent do you believe you should reduce the amount of alcohol you consume"—could be interpreted as measuring normative perceptions, rather than intentions. Future research should use more standardized intentions measures with established reliability and validity (Francis et al., 2004), should measure drinking intentions before and after message exposure to quantify changes in intentions (Pettigrew, Jongenelis, et al., 2016) or should measure actual drinking behaviours.

In Study 2, credibility questions preceded intentions questions, which may have primed participants to critically evaluate the message with rational System 2 thinking, thereby diminishing any impact on intentions via intuitive System 1 thinking (Kahneman, 2011). Future research could present intentions questions before credibility questions. Also, the moderation analyses were exploratory. Future research should pre-register the moderation hypotheses that emerged from this study and be powered to detect moderation effects, which this study was not.

A strength is the novel exploration of AI funding disclosure on credibility and drinking intentions and prototype perceptions as moderators.

Implications

Alcohol SAPROs disseminate information with the stated objective of reducing harmful drinking (Drinkaware, n.d; Drinkwise, n.d.a). However, exploratory analyses in Study 1 suggest it could be hypothesised that some messages used in such campaigns- specifically those presenting an unhealthy norm and the benefits of drinking- may adversely impact certain crucial groups, namely those identifying more as heavy drinkers and less as responsible drinkers. Although the messages examined did not comprehensively reflect those used in SAPRO campaigns, this hypothesis echoes concerns about possible adverse impacts of AI-funded health campaigns (Babor et al., 2018; McCambridge et al., 2014; Petticrew et al., 2020).

Study 2 has implications for the question of whether AI funding should be disclosed in SAPRO campaigns. Given that source credibility was perceived as significantly lower when obfuscating messages were combined with funding disclosure, an obligation to disclose funding could serve as a safeguard that both prevents consumers being misled and motivates SAPROs to avoid obfuscating messages altogether. This is important given widespread misperceptions about these groups' funding (Brennan et al., 2017; Pettigrew, Biagioni, et al., 2016) and findings that exposure to AI-funded messages increases uncertainty and false certainty about the links between alcohol and breast cancer (Maani et al., 2022). In the clear, factual single causality conditions, funding disclosure made no difference to perceived

credibility and enhanced these messages' impact on drinking intentions among those identifying as more similar to heavy drinkers and less similar to responsible drinkers. This suggests that when providing clear, factual information, SAPROs would not lose credibility from disclosure messages and may indeed enhance their campaigns' impact among these important groups.

CONCLUSION

Taken together, these studies contribute to the growing body of research exploring whether AI-funded responsible drinking campaigns are at best suboptimal, and at worst, harmful (Brennan et al., 2020; Maani et al., 2022; Moss et al., 2015; Petticrew et al., 2020). The present studies did not find that social norm, framing, alternative causation or funding disclosure manipulations presented in an artificial experimental context significantly affected drinking intentions. However, Study 2 found that disclosure of industry funding significantly affected judgements of source credibility in messages about the links between alcohol and cancer. The exploratory moderation analyses provide hypotheses regarding the type of people who are most vulnerable to dark nudges in AI-funded campaigns. The exploratory moderation analyses suggest prototype perceptions merit further research as moderators of alcohol health interventions.

AUTHOR CONTRIBUTIONS

Joel Lewin: Conceptualization; formal analysis; investigation; methodology; visualization; writing – original draft; writing – review and editing. **Matt Field:** Supervision; writing – review and editing. **Emma Davies:** Conceptualization; methodology; supervision; writing – review and editing.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study will be openly available at http://doi.org/10.17605/OSF.IO/YUN37.

ORCID

Joel Lewin https://orcid.org/0000-0003-3688-4920
Matt Field https://orcid.org/0000-0002-7790-5559
Emma Davies https://orcid.org/0000-0003-3577-3276

REFERENCES

Atkin, J. L., Mccardle, M., & Newell, S. J. (2008). The role of advertiser motives in consumer evaluations of 'responsibility' messages from the alcohol industry. *Journal of Marketing Communications*, *14*(4), 315–335. https://doi.org/10.1080/13527260802141447

Babor, T. F. (2009). Alcohol research and the alcoholic beverage industry: Issues, concerns and conflicts of interest. *Addiction*, *104*, 34–47. https://doi.org/10.1111/j.1360-0443.2008.02433.x

Babor, T. F., Robaina, K., Brown, K., Noel, J., Cremonte, M., Pantani, D., Peltzer, R. I., & Pinsky, I. (2018). Is the alcohol industry doing well by 'doing good'? Findings from a content analysis of the alcohol industry's actions to reduce harmful drinking. *BMJ Open*, *8*(10), e024325. https://doi.org/10.1136/bmjopen-2018-024325

Blackwell, A. K. M., Clarke, N., Pechey, E., & Attwood, A. S. (2021). Alcohol labelling: Evidence for product information interventions. In R. Cooke, D. Conroy, E. L. Davies, M. S. Hagger, & R. O. de Visser (Eds.), *The Palgrave handbook of psychological perspectives on alcohol consumption* (pp. 451–475). Palgrave Macmillan. https://doi.org/10.1007/978-3-030-66941-6_19

Blackwell, A. K. M., Drax, K., Attwood, A. S., Munafò, M. R., & Maynard, O. M. (2018). Informing drinkers: Can current UK alcohol labels be improved? *Drug and Alcohol Dependence*, *192*, 163–170. https://doi.org/10.1016/j.drugalcdep.2018.07.032

Brennan, E., Schoenaker, D. A. J. M., Durkin, S. J., Dunstone, K., Dixon, H. G., Slater, M. D., Pettigrew, S., & Wakefield, M. A. (2020). Comparing responses to public health and industry-funded alcohol harm reduction advertisements: An experimental study. *BMJ Open*, *10*(9), e035569. https://doi.org/10.1136/bmjopen-2019-035569

Brennan, E., Wakefield, M. A., Durkin, S. J., Jernigan, D. H., Dixon, H. G., & Pettigrew, S. (2017). Public awareness and misunderstanding about DrinkWise Australia: A cross-sectional survey of Australian adults. *Australian and New Zealand Journal of Public Health*, *41*(4), 352–357. https://doi.org/10.1111/1753-6405.12674

Byrne, S., Guillory, J. E., Mathios, A. D., Avery, R. J., & Hart, P. S. (2012). The unintended consequences of disclosure: Effect of manipulating sponsor identification on the perceived credibility and effectiveness of smoking cessation advertisements. *Journal of Health Communication*, *17*(10), 1119–1137. https://doi.org/10.1080/10810730.2012.665425

Churchill, S., Pavey, L., Jessop, D., & Sparks, P. (2016). Persuading people to drink less alcohol: The role of message framing, temporal focus and autonomy. *Alcohol and Alcoholism*, *51*(6), 727–733. https://doi.org/10.1093/alcalc/agw033

Davies, E. L. (2019). Similarity to prototypical heavy drinkers and non-drinkers predicts AUDIT-C and risky drinking in young adults: Prospective study. *Psychology & Health*, *34*(4), 403–421. https://doi.org/10.1080/08870446.2018.1532510

Davies, E. L., Lewin, J., & Field, M. (2022). Am I a responsible drinker? The impact of message frame and drinker prototypes on perceptions of alcohol product information labels. *Psychology & Health*, *1–18*, 1–18. https://doi.org/10.1080/08870446.2022.2129055

Davies, E. L., & Todd, J. (2021). Drinking beyond intentions: The prototype willingness model and alcohol consumption. In R. Cooke, D. Conroy, E. L. Davies, M. S. Hagger, & R. O. de Visser (Eds.), *The Palgrave handbook of psychological perspectives on alcohol consumption* (pp. 51–76). Palgrave Macmillan. https://doi.org/10.1007/978-3-030-66941-6_3

Dimova, E. D., & Mitchell, D. (2022). Rapid literature review on the impact of health messaging and product information on alcohol labelling. *Drugs: Education, Prevention and Policy*, *29*(5), 451–463. https://doi.org/10.1080/09687637.2021.1932754

Drinkaware. (n.d.). What we do. https://www.drinkaware.co.uk/about-us/what-we-do

Drinkaware [@drinkaware]. (2020). *Our new research out today reveals substantial patterns in the nation's drinking over the lockdown. One in 10 drinkers – equivalent…* [Tweet: image attached]. Twitter. https://twitter.com/Drinkaware/status/1323310749725458435?s=20

Drinkaware IRL [@Drinkaware_ie]. (2020). *Drinking alcohol is an unhelpful coping strategy for #COVID19 stress or anxiety. In these extraordinary times, let's find healthier and….* [Tweet; image attached]. https://twitter.com/Drinkaware_ie/status/1246020213440753667?s=20

Drinkwise. (n.d.-a). About. https://drinkwise.org.au/about-us/about/#

Drinkwise. (n.d.-b). Is your drinking putting you at risk of cancer? https://drinkwise.org.au/drinking-and-you/is-your-drinking-putting-you-at-risk-of-cancer/#

Eagly, A. H., & Chaiken, S. (1975). An attribution analysis of the effect of communicator characteristics on opinion change: The case of communicator attractiveness. *Journal of Personality and Social Psychology*, *32*(1), 136–144. https://doi.org/10.1037/h0076850

Éduc'alcool. (2017). The 8 benefits of moderate drinking. https://www.educalcool.qc.ca/en/facts-and-consequences/alcohol-and-health/8-benefits-of-moderate-drinking/

Foxcroft, D. R., Moreira, M. T., Almeida, S. N. M., & Smith, L. A. (2015). Social norms information for alcohol misuse in university and college students. *The Cochrane Database of Systematic Reviews*, *12*(12), 006748. https://doi.org/10.1002/14651858.CD006748.pub4

Francis, J., Eccles, M., Johnston, M., Walker, A., Grimshaw, J., Foy, R., Kaner, E. F. S., Smith, L., & Bonetti, D. (2004). *Theory of planned behaviour questionnaires: Manual for researchers*. Centre for Health Services Research, University of Newcastle.

Harris, P. R., Sillence, E., & Briggs, P. (2009). The effect of credibility-related design cues on responses to a web-based message about the breast cancer risks from alcohol: Randomized controlled trial. *Journal of Medical Internet Research*, *11*(3), e37. https://doi.org/10.2196/jmir.1097

Hassan, L. M., & Shiu, E. (2018). A systematic review of the efficacy of alcohol warning labels: Insights from qualitative and quantitative research in the new millennium. *Journal of Social Marketing*, *8*(3), 333–352. https://doi.org/10.1108/JSOCM-03-2017-0020

Hayes, A. F. (2018). *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach* (2nd ed.). The Guilford Press.

Hobin, E., Weerasinghe, A., Vallance, K., Hammond, D., McGavock, J., Greenfield, T. K., Schoueri-Mychasiw, N., Paradis, C., & Stockwell, T. (2020). Testing alcohol labels as a tool to communicate cancer risk to drinkers: A real-world quasi-experimental study. *Journal of Studies on Alcohol and Drugs*, *81*(2), 249–261.

Kahneman, D. (2011). *Thinking, fast and slow*. Farrar.

Kingsbury, J. H., Gibbons, F. X., & Gerrard, M. (2015). The effects of social and health consequence framing on heavy drinking intentions among college students. *British Journal of Health Psychology*, *20*(1), 212–220. https://doi.org/10.1111/bjhp.12100

Maani, N., van Schalkwyk, M. C. I., Filippidis, F. T., Knai, C., & Petticrew, M. (2022). Manufacturing doubt: Assessing the effects of independent vs industry-sponsored messaging about the harms of fossil fuels, smoking, alcohol, and sugar sweetened beverages. *SSM – Population Health*, *17*, 101009. https://doi.org/10.1016/j.ssmph.2021.101009

McCambridge, J., Kypri, K., Miller, P., Hawkins, B., & Hastings, G. (2014). Be aware of Drinkaware. *Addiction*, *109*(4), 519–524. https://doi.org/10.1111/add.12356

Moss, A. C., Albery, I. P., Dyer, K. R., Frings, D., Humphreys, K., Inkelaar, T., Harding, E., & Speller, A. (2015). The effects of responsible drinking messages on attentional allocation and drinking behaviour. *Addictive Behaviors*, *44*, 94–101. https://doi.org/10.1016/j.addbeh.2014.11.035

Nakamura, R., Pechey, R., Suhrcke, M., Jebb, S. A., & Marteau, T. M. (2014). Sales impact of displaying alcoholic and non-alcoholic beverages in end-of-aisle locations: An observational study. *Social Science & Medicine*, *108*, 68–73. https://doi.org/10.1016/j.socscimed.2014.02.032

Newall, P. W. S. (2013). Dark Nudges. https://darksideofthenudge.wordpress.com/2013/04/19/dark-nudges/

Newall, P. W. S. (2019). Dark nudges in gambling. *Addiction Research & Theory*, *27*(2), 65–67. https://doi.org/10.1080/16066359.2018.1474206

Newall, P. W. S., Weiss-Cohen, L., Singmann, H., Walasek, L., & Ludvig, E. A. (2022). Impact of the "when the fun stops, stop" gambling message on online gambling behaviour: A randomised, online experimental study. *The Lancet Public Health.*, *7*, e437–e446. https://doi.org/10.1016/S2468-2667(21)00279-6

Park, S. Y., Son, H., Lee, J., & Go, E. (2020). Moderating effects of social norms and alcohol consumption on message framing in responsible drinking campaigns: Value from deviance regulation theory. *Health Communication*, *35*(7), 793–803. https://doi.org/10.1080/10410236.2019.1593077

Pennington, C. R., Monk, R. L., Qureshi, A. W., Kulkarni, R., Li, W., Li, J., & Heim, D. (2022). Dark nudges: Branding magnifies the decoy effect in alcohol purchasing decisions. *Journal of Studies on Alcohol and Drugs*, *83*(6), 857–866. https://doi.org/10.15288/jsad.21-00398

Petticrew, M., Fitzgerald, N., Durand, M. A., Knai, C., Davoren, M., Perry, I., & Vrana, K. E. (2016). Diageo's 'stop out of control drinking' campaign in Ireland: An analysis. *PLoS One*, *11*(9), 0160379. https://doi.org/10.1371/journal.pone.0160379

Petticrew, M., Maani Hessari, N., Knai, C., & Weiderpass, E. (2018). How alcohol industry organisations mislead the public about alcohol and cancer. *Drug and Alcohol Review*, *37*(3), 293–303. https://doi.org/10.1111/dar.12596

Petticrew, M., Maani, N., Pettigrew, L., Rutter, H., & Van Schalkwyk, M. C. (2020). Dark nudges and sludge in big alcohol: Behavioral economics, cognitive biases, and alcohol industry corporate social responsibility. *The Milbank Quarterly*, *98*(4), 1290–1328. https://doi.org/10.1111/1468-0009.12475

Pettigrew, S., Biagioni, N., Daube, M., Stafford, J., Jones, S. C., & Chikritzhs, T. (2016). Reverse engineering a 'responsible drinking' campaign to assess strategic intent. *Addiction*, *111*(6), 1107–1113. https://doi.org/10.1111/add.13296

Pettigrew, S., Jongenelis, M. I., Glance, D., Chikritzhs, T., Pratt, I. S., Slevin, T., Liang, W., & Wakefield, M. (2016). The effect of cancer warning statements on alcohol consumption intentions. *Health Education Research*, *31*(1), 60–69. https://doi.org/10.1093/her/cyv067

Reinert, D. F., & Allen, J. P. (2007). The alcohol use disorders identification test: An update of research findings. *Alcoholism: Clinical and Experimental Research*, *31*(2), 185–199. https://doi.org/10.1111/j.1530-0277.2006.00295.x

Sim, F., Chick, J., Jarvis, S., Lidington, I., Neidle, S., Ogden, G., & Swerdlow, F. (2018). Rapid response to: Public Health England's capture by the alcohol industry. *BMJ*, *2018*, 362. https://doi.org/10.1136/bmj.k3928

Smith, C. T., De Houwer, J., & Nosek, B. A. (2013). Consider the source. *Personality and Social Psychology Bulletin*, *39*(2), 193–205. https://doi.org/10.1177/0146167212472374

Smith, S. W., Atkin, C. K., & Roznowski, J. (2006). Are "drink responsibly" alcohol campaigns strategically ambiguous? *Health Communication*, *20*(1), 1–11. https://doi.org/10.1207/s15327027hc2001_1

Sparkman, R. M., Jr. (1982). The discounting principle in the perception of advertising. *Advances in Consumer Research*, *9*, 277–280. https://www.acrwebsite.org/volumes/6008/volumes/v09/NA-09

Stockwell, T., Solomon, R., O'Brien, P., Vallance, K., & Hobin, E. (2020). Cancer warning labels on alcohol containers: A consumer's right to know, a Government's responsibility to inform, and an Industry's power to thwart. *Journal of Studies on Alcohol and Drugs*, *81*(2), 284–292. https://doi.org/10.15288/jsad.2020.81.284

Swann, W. B., & Read, S. J. (1981). Self-verification processes: How we sustain our self-conceptions. *Journal of Experimental Social Psychology*, *17*(4), 351–372. https://doi.org/10.1016/0022-1031(81)90043-3

Szykman, L. R., Bloom, P. N., & Blazing, J. (2004). Does corporate sponsorship of a socially-oriented message make a difference? An investigation of the effects of sponsorship identity on responses to an anti-drinking and driving message. *Journal of Consumer Psychology*, *14*(1–2), 13–20. https://doi.org/10.1207/s15327663jcp1401&2_3

Thaler, R. H., & Sunstein, C. R. (2008). *Nudge: Improving decisions about health, wealth, and happiness*. Yale University Press.

Walster, E., Aronson, E., & Abrahams, D. (1966). On increasing the persuasiveness of a low prestige communicator. *Journal of Experimental Social Psychology*, *2*(4), 325–342. https://doi.org/10.1016/0022-1031(66)90026-6

Webb, T. L., & Sheeran, P. (2006). Does changing behavioral intentions engender behavior change? A meta-analysis of the experimental evidence. *Psychological Bulletin*, *132*(2), 249–268. https://doi.org/10.1037/0033-2909.132.2.249

Wells, G. L., & Windschitl, P. D. (1999). Stimulus sampling and social psychological experimentation. *Personality and Social Psychology Bulletin*, *25*(9), 1115–1125. https://doi.org/10.1177/01461672992512005

Winstock, A. R., Holmes, J., Ferris, J. A., & Davies, E. L. (2020). Perceptions of alcohol health warning labels in a large international cross-sectional survey of people who drink alcohol. *Alcohol and Alcoholism*, *55*(3), 315–322. https://doi.org/10.1093/alcalc/agz099

World Health Organization. (2016). Global NCD Target: Reducing Harmful Use Of Alcohol. https://www.who.int/beat-ncds/take-action/ncd-brief-alcohol.pdf?ua=1#:~:text=use%20of%20alcohol.-,Global%20Target,the%20national%20context%20by%202025

World Health Organization. (2018). Global status report on alcohol and health 2018: Executive summary. https://apps.who.int/iris/rest/bitstreams/1217404/retrieve

Zhao, J., Stockwell, T., Vallance, K., & Hobin, E. (2020). The effects of alcohol warning labels on population alcohol consumption: An interrupted time series analysis of alcohol sales in Yukon, Canada. *Journal of Studies on Alcohol and Drugs*, *81*(2), 225–237. https://doi.org/10.15288/jsad.2020.81.225

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Lewin, J., Field, M., & Davies, E. (2023). Investigating the impact of 'dark nudges' on drinking intentions: A between groups, randomized and online experimental study. *British Journal of Health Psychology*, *00*, 1–21. https://doi.org/10.1111/bjhp.12698