# scientific reports

Check for updates

## **OPEN** Comparing attentional disengagement between Prolific and MTurk samples

Derek A. Albert<sup>⊠</sup> & Daniel Smilek

Attention often disengages from primary tasks in favor of secondary tasks (i.e., multitasking) and task-unrelated thoughts (i.e., mind wandering). We assessed whether attentional disengagement, in the context of a cognitive task, can substantially differ between samples from commonly used online participant recruitment platforms, Prolific and Mechanical Turk (MTurk). Initially, eighty participants were recruited through Prolific to perform an attention task in which the risk of losing points for errors was varied (high risk = 80% chance of loss, low risk = 20% chance of loss). Attentional disengagement was measured via task performance along with self-reported mind wandering and multitasking. On Prolific, we observed surprisingly low levels of disengagement. We then conducted the same experiment on MTurk. Strikingly, MTurk participants exhibited more disengagement than Prolific participants. There was also an interaction between risk and platform, with the high-risk group exhibiting less disengagement, in terms of better task performance, than the low-risk group, but only on MTurk. Platform differences in individual traits related to disengagement and relations among study variables were also observed. Platform differences persisted, but were smaller, after increasing MTurk reputation criteria and remuneration in a second experiment. Therefore, recruitment platform and recruitment criteria could impact results related to attentional disengagement.

Attention is seldom fully engaged in the task at hand, since it often disengages in favor of secondary tasks, such as consuming media (i.e., media multitasking) or entertaining task-unrelated thoughts (i.e., mind wandering). The degree and type of disengagement can vary drastically between individuals and contexts, however. For instance, disengagement is known to vary as a function of individual traits (e.g., personality)<sup>1,2</sup> and population (e.g., young vs. old)<sup>3,4</sup>, as well as task properties (e.g., difficulty)<sup>5,6</sup> and environment (e.g., laboratory vs. everyday life)<sup>7,8</sup>. One factor that has become increasingly important in psychological research relates to the emergence of online participant recruitment platforms. The widespread use of these platforms has sparked concerns over the potential increase in opportunities for disengagement during online compared to in-person testing. Indeed, Drody and colleagues9 found that, across 16 online studies, approximately 38% (range 9-85%) of participants admitted to engaging in unrelated tasks during testing. While findings from studies comparing attention and data quality between in-person and online contexts are mixed<sup>10,11</sup>, early attempts to replicate well-established findings by administering traditionally lab-based cognitive tasks online has proven largely successful<sup>12</sup>. Complicating matters, however, is the current availability of multiple online recruitment platforms, such as Amazon's Mechanical Turk (MTurk) and the more recently introduced Prolific<sup>13</sup>. Despite widespread use of these platforms, it is unclear how they compare in terms of disengagement in cognitive tasks. Here, we report substantial differences between MTurk and Prolific in terms of participant disengagement and performance on an attention task.

Our interest in comparing disengagement and performance between competing online platforms emerged from a quite different initial objective, which was to examine whether attentional disengagement can be influenced by variations in the risk of negative consequences for poor performance. This objective relates to a large body of work that has documented the detrimental effect of disengagement on task performance. For instance, media multitasking is associated with poor memory encoding, reading comprehension, and information recall in educational settings<sup>14,15</sup>. Additionally, mind wandering is associated with slow and variable stimulus reaction times<sup>16,17</sup>, failure to inhibit automatic responses<sup>18,19</sup>, and impaired visual search<sup>20</sup>. In certain contexts, these performance deficits can have negative consequences, ranging from poor grades in school<sup>21,22</sup> to fatalities from accidents in the workplace<sup>23</sup> or on the road<sup>24</sup>. Mind wandering and media multitasking can offer certain benefits, such as boosting creativity<sup>25,26</sup>, facilitating prospection<sup>27,28</sup>, and maintaining task motivation<sup>29</sup>, but given the aforementioned findings, it is likely adaptive to decrease task disengagement in circumstances when the risk of negative consequences for it is high.

Department of Psychology, University of Waterloo, Waterloo, Ontario, Canada. 🖾 email: dalbert@uwaterloo.ca

There is already some evidence consistent with the hypothesis that disengagement from a primary task decreases when the risk of negative consequences from disengagement increases. For example, findings from traffic safety research suggest that greater crash risk may motivate drivers to reduce disengagement. Drivers who perceive the risk of a crash associated with texting, talking to passengers, and mind wandering to be high also report less of these forms of disengagement <sup>30,31</sup>. However, while these findings are consistent with the notion that individuals decrease disengagement when the perceived risk of negative consequences is high, they may alternatively be explained by individual traits, including personality (e.g., sensation seeking)<sup>32,33</sup> and cognitive capacities (e.g., executive functioning)<sup>34,35</sup>. Another finding consistent with the idea that increased risk of negative consequences is linked to less disengagement is that drivers report less mind wandering while navigating sections of a familiar route with high versus low crash rates<sup>36</sup>. However, this finding too has a possible alternative explanation, namely that areas of increased crash risk involve more challenging (i.e., difficult) driving conditions, and it is the increased difficulty (and not specifically the increased risk) that reduces disengagement in those stretches of road. Thus, while suggestive, these prior studies provide insufficient evidence to clarify whether risk of negative consequences is causally linked to disengagement.

In the present study, we initially sought to build on previous work by directly manipulating risk or the probability of negative consequences (losing points) for poor performance (incorrect responses) in the context of an attention task. In so doing, we aimed to examine the influence of risk on disengagement in terms of mind wandering, multitasking, and performance, while holding task difficulty constant. We recruited participants through Prolific and had them perform a cognitive task that demanded sustained attention, over the course of which they accumulated points, but risked losing points for incorrect responses. Participants were randomly assigned to either a high-risk condition, in which the likelihood of losing points for an error was 80%, or a lowrisk condition, in which the likelihood of losing points for an error was 20%. Participants were told that the attention task would end as soon as they reached a certain level of performance. To assess attentional disengagement, we periodically asked participants whether or not they were mind wandering and, at the end of the study, whether or not they were multitasking (media-related or media-unrelated) during the attention task. Finally, we also measured individual traits related to risk tolerance (sensation seeking) and attentional disengagement in everyday life (inattention and cognitive errors) to establish whether they positively relate to one another.

However, after examining the initial data collected through Prolific (see Fig. 3), we added another research focus. Analyses of the Prolific sample revealed rates of self-reported mind wandering and media multitasking that were markedly lower than in studies with MTurk samples and other tasks<sup>9,37</sup>. Given the novelty of the attention task paradigm that we used to manipulate risk in the present study, it was unclear whether the low levels of disengagement in our Prolific sample were attributable to the platform, or whether they were linked to the unique properties of the task. As a result, we became interested to know whether experimental results would differ between samples recruited through Prolific and MTurk while the task was held constant across platforms. Accordingly, in addition to studying the impact of risk on disengagement, we set out to directly compare results between platforms by conducting the same study we implemented on Prolific, using MTurk.

Numerous studies report attentional disengagement differences between online recruitment platforms in the context of surveys, where disengagement has primarily been assessed using attention-check questions<sup>38–41</sup>. This technique relies on trick questions (e.g., "Have you ever had a fatal heart attack?") and question-specific instructions (e.g., "Ignore the question asking which sport you like.") to detect when participants have not attended to survey content. Findings from studies comparing recruitment platforms using this method are mixed. Some studies report less disengagement in Prolific samples than MTurk samples<sup>38,41</sup>, while others report approximately equal or less disengagement in MTurk samples<sup>40,42</sup>. Various factors may account for this variability, including prior experience with attention-check questions, which tends to be higher in MTurk samples<sup>43</sup>. Another factor is reputation criteria, including human intelligence task (HIT) approval rating and number of completed HITs, which are commonly used in MTurk studies<sup>44,45</sup>. However, some evidence suggests that increasing reputation criteria does little to reduce disengagement<sup>46</sup>, which is why services like CloudResearch offer additional controls for MTurk<sup>41,42,47</sup>. There is a growing number of studies using online recruitment platforms, and native MTurk in particular, to collect cognitive task data<sup>48-54</sup>. Given that most between-platform comparisons have focused on disengagement assessed using attention-check questions in the context of surveys, it is unclear to what extent findings from these comparisons generalize to cognitive tasks. Accordingly, the present study compared performance on an attention task, subjective reports of disengagement, and the effects of a risk manipulation on these variables between two online recruitment platforms, namely Prolific and native MTurk.

In summary, the present study had two aims: (1) examine the influence of risk on attentional disengagement, and (2) examine potential differences in attentional disengagement and experimental outcomes between online recruitment platforms. In an initial experiment (Experiment 1 or E1), we pursued the first aim by manipulating the risk or probability of losing points for incorrect responses on an attention task and randomly, equally (1:1) allocating participants to either a low-risk condition (20% chance of losing points) or a high-risk condition (80% chance of losing points). We hypothesized that increasing risk would lead participants to perform better on the task and decrease disengagement, in terms of mind wandering and multitasking, despite not being informed about the level of risk. We also tested for positive correlations between individual traits related to risk tolerance and attentional disengagement. We pursued the second aim in E1 by collecting data from MTurk in addition to Prolific. Given the surprisingly low levels of disengagement we initially observed for Prolific, we hypothesized that disengagement would be higher for MTurk. We also assessed whether the effects of our risk manipulation differed between platforms. Furthermore, we compared Prolific and MTurk in terms of relations between study variables.

Experiment 2 (E2) addressed two issues that emerged after E1 was completed. First, following initial submission of our manuscript, Rioja and colleagues published a study showing no significant difference in performance on a brief attention task between Prolific and MTurk<sup>55</sup>. In contrast to our E1—which used MTurk reputation criteria based on those used in previous studies comparing survey data quality between platforms<sup>56,57</sup>—Rioja and

colleagues used stricter reputation criteria on MTurk, suggesting that increasing MTurk criteria may decrease platform differences. Accordingly, in E2, we replicated the reputation criteria used by Rioja and colleagues and compared the results to those of E1. Second, during peer review, a concern about the equivalence of remuneration between platforms in E1 emerged. In E1, we determined the USD amount offered to MTurk participants by converting the GBP amount offered to Prolific participants using the exchange rate at the time of obtaining ethical approval for the study. However, it became clear that the difference in purchasing power between currencies may influence participant self-selection and motivation. Thus, in E2, we increased MTurk remuneration to match Prolific remuneration in terms of purchasing power parity, which accounts for both the exchange rate and difference in cost of goods between two currencies<sup>58</sup>.

### Methods

#### **Participants**

All study procedures were approved by the University of Waterloo Office of Research Ethics (ORE #43709) and conducted in accordance with the Declaration of Helsinki. Informed consent was obtained from all participants.

In E1, target sample sizes of n = 80 participants were recruited from both Prolific and MTurk (N=160). Inclusion and exclusion criteria were assessed using a screening questionnaire hosted on Qualtrics. To participate, candidates had to (1) be in the United States or Great Britain, (2) self-report an age between 18 and 35 years, (3) self-report normal or corrected visual acuity, (4) self-report normal colour vision, and (5) self-report no mental health issues. MTurk candidates were also required to have completed more than 100 HITs and have a HIT approval rating greater than 95%. These parameters were chosen based on recommendations<sup>45</sup> and findings from studies that used survey-based attention checks to examine data quality as a function of MTurk worker reputation (i.e., approval rating) and productivity (i.e., HITs completed)<sup>36,57</sup>, as well as recruitment platform<sup>39</sup>. Candidates who did not meet criteria, or who did not consent to participate in the study, were not permitted to continue and asked to withdraw using the designated method on their recruitment platform. The study took approximately 45 min to complete, and participants were remunerated £5.75 GBP on Prolific and \$6.25 USD on MTurk, which were roughly equivalent given exchange rates at the time. Recruitment and testing for E1 took place over one month (November 2022).

E2 recruitment and study procedures were identical to those of E1, but with some key exceptions. In contrast to E1, Prolific and MTurk candidates in E2 were both required to have an approval rating greater than 95%. Furthermore, MTurk candidates were required to have completed more than 1000 HITs. There parameters were chosen to replicate the recent study by Rioja and colleagues that found no significant difference in performance on a brief attention task between Prolific and MTurk<sup>55</sup>. Additionally, remuneration for MTurk participants was increased to \$8.40 USD in E2, which was equivalent to the £5.75 GBP offered to Prolific participants after accounting for both exchange rate and the average cost of goods in each currency (i.e., purchasing power parity)<sup>58</sup>. Recruitment and testing for E2 took place over two weeks (late July 2023).

#### Measures

#### Demographics questionnaire

Selection and wording of demographic variables were based on those used by Statistics Canada to collect census data. Participants were asked, but could decline, to self-report the following demographic variables: date of birth (month, day, year), which was used to calculate age; sex at birth (male, female, intersex); gender (man, woman, non-binary), which could differ from sex at birth and on legal documents; current education/employment status (e.g., full-time work, full-time studies, part-time work + full-time studies, caregiver); current occupation (if applicable); and highest level of education completed (e.g., high school, some college, some university).

#### Individual trait questionnaires

The Attention-Related Cognitive Errors Scale (ARCES) measured the trait tendency to experience cognitive errors from attentional disengagement in everyday life. The ARCES contains 12 items and has a single-factor structure. Participants respond to statements (e.g., "I begin one task and get distracted into doing something else.") by indicating how frequently they experience each using a five-point Likert scale (1 = "Never" to 5 = "Very often"). The scale shows good internal consistency with a Cronbach's alpha of  $0.88^{59}$ .

The Mindful Attention Awareness Scale – Lapses Only (MAAS) assessed the trait tendency to experience attentional lapses, or inattention resulting from disengagement, in everyday life. The MAAS-LO is a subset of the MAAS<sup>60</sup> and contains 12 items that specifically relate to attentional lapses. Participants respond to statements (e.g., "It seems I am 'running on automatic' without much awareness of what I am doing.") by indicating how frequently they experience each using a six-point Likert scale (1="Almost never" to 6="Almost always"). The scale shows good internal consistency with a Cronbach's alpha of  $0.83^{61}$ .

The Workplace Cognitive Failure Scale (WCFS) measured the trait tendency to experience cognitive errors from attentional disengagement at work. The WCFS contains 15 items and has a three-factor structure relating to failures in memory, attention, and motor action<sup>23</sup>. Participants respond to questions (e.g., memory: "Failure to recall work procedures?", attention: "Daydream when you ought to be listening to somebody?", motor action: "Accidently drop objects or things?") by indicating how frequently they experience each using a five-point Likert scale (1 = "Never" to 5 = "Very often").

The Brief Sensation Seeking Scale (BSSS) measured the personality trait of sensation seeking or risk tolerance. It contains eight items and has a single-factor structure. Participants response to statements (e.g., "I like to do frightening things.") by indicating the extent to which they agree with each using a five-point Likert scale (1 = "Strongly disagree" to 5 = "Strongly agree"). The scale shows good internal consistency with a Cronbach's alpha of  $0.76^{62}$ .

#### Multitasking questionnaire

Participants were asked whether they multitasked during the study with the following question from Drody and colleagues<sup>51</sup>: "While completing this study, were you engaged in any media-related activities outside of the contents of the experiment (e.g., attending to content in another browser, listening to music or using a smartphone/ tablet while completing the study?" Response options included: (1) "Yes"; (2) "No, I did not engage in any activities outside of the contents of this study"; and (3) "No, but I was engaged in other, media-unrelated activities while completing this study." If participants indicated 1 (media-related multitasking) or 3 (media-unrelated multitasking), they were asked two follow-up questions: (1) "How much of your attention was directed to media-related or media-unrelated activities outside of the contents of the experiment was affected by your media-related or media-unrelated activities?" Likert response options ranged from 1 = "None at all" to 5 = "A great deal." Participants were asked to answer honestly with the assurance that they would not be penalized for responding one way or another.

#### Attention task

Figure 1 illustrates the attention task with performance feedback and risk manipulation. The attention task consisted of a 1-back task adapted from Konishi and colleagues<sup>63</sup>. Specifically, our adaptation was designed to be more difficult than the original to avoid ceiling performance effects, which might obscure differences between risk conditions and/or recruitment platforms. The task included non-target and target displays. Non-target displays consisted of four shapes arranged randomly in a  $2 \times 2$  grid, with vertical and horizontal lines delineating the cells. There were six possible shapes including a circle, triangle, square, star, diamond, and pentagon. Only one of each shape was displayed at any given time, resulting in 360 possible arrangements. Target displays consisted of one shape, centred on the display, that was randomly selected from the four displayed in the previous, non-target display. In addition, four question marks were arranged in a  $2 \times 2$  grid around the middle shape. The shapes were presented in blue on a grey background.

Each trial included two to six non-target displays presented sequentially, followed by a single target display. Non-targets were displayed for 1300–1700 ms and targets were displayed for 2100–2500 ms. The interval between non-target and target displays, during which no shapes were present, ranged from 2000 to 4000 ms. All the timings varied in 100 ms increments. On target displays, a shape was presented in the centre and participants indicated which cell contained the shape in the last non-target display by pressing a corresponding key on the keyboard. There were 80 trials in total to ensure the task was long enough to observe vigilance decrements, both in terms of decreased performance and increased mind wandering, which may distinguish conditions and platforms. Furthermore, we anticipated that participants would require several trials to learn the probability of losing points from incorrect responses.

Performance feedback was delivered after each target. Correct responses and incorrect responses were signalled by the words "Correct" and "Incorrect," respectively. Additionally, participants saw two values: one labelled "Points" and another labelled "Streaks." As participants progressed through the task, they accumulated points. For every 5 points, they were awarded a streak (and their points reset to zero). If participants responded incorrectly, there was a chance they would lose all of their currently accumulated points. Risk was manipulated by varying



**Figure 1.** Attention task with performance feedback and risk manipulation. There were two to six non-target displays before a target display. There was a 20% chance of a mind wandering probe preceding performance feedback. The "x" on the line between the probe and bottom right performance feedback displays represents the chance of participants losing points for an incorrect response. The "100%-x" on the line between the probe and bottom left performance feedback displays represents the chance of not losing points for an incorrect response.

the probability of losing points as a result of an incorrect response. The probability of losing points was 20% for the low-risk condition, and 80% for the high-risk condition. Participants were told that the task would end as soon as they achieved, "a certain level of performance," although no such criterion existed.

Attention task performance was assessed using an accuracy score, calculated as the proportion of correct responses over the total number of target trials. Thought probes, delivered during the attention task, measured mind wandering. Probes appeared immediately following target trials and before performance feedback. There was a 20% chance of a probe appearing after each target trial. The probes asked participants: "Were you mind wandering just before seeing the question marks?" Mind wandering was defined in the initial task instructions as "having thoughts unrelated to the task" and the question marks referred to the target display. Participants used their keyboard to select between "Yes" and "No" choice options. Mind wandering was calculated as the proportion of "Yes" responses over the total number of probes. To assess perception of risk, participants were asked the following question at the end of the cognitive task: "How likely do you think it was that an incorrect response in the task would lead to a loss of points?" They inputted their response using a continuous slider between 0% and 100%, which was coded as a proportion ranging from 0 to 1.

#### Procedure

Study procedures were identical for Prolific and MTurk samples. After registering for the study on Prolific or MTurk, study candidates followed a link to a Qualtrics survey where they completed a screening questionnaire. Then, if eligible, participants were shown an information and consent form. After providing informed consent, participants filled out the demographics questionnaire followed by the ARCES, MAAS, WCFS, and BSSS, which took approximately 15 min. Participants were then redirected to a webpage with the attention task, at which point they were randomly assigned to either the low-risk condition or high-risk condition. After reading through the task instructions, participants completed the task, which took approximately 25 min. Participants were then redirected back to Qualtrics where they completed the multitasking questionnaire and were fully debriefed on the purposes of the study.

#### Analyses

Identical analyses were performed for E1 and E2 using R statistical software (version 4.2.3). We compared demographic characteristics between platforms using a t-test for age and chi-squared tests for sex, gender, education, and employment. We also compared dropout rates and study completion times between platforms using chisquared tests and t-tests, respectively.

Our main interest was to examine the effects of condition, platform, and their interaction on task accuracy, mind wandering, and risk perception. For accuracy and mind wandering, general linear mixed models (GLMMs) were generated using packages lme4 (version 1.1-32) and afex (version 1.2-1). Since these variables are binary (correct/incorrect and focused/mind wandering) and were measured repeatedly (via multiple trials/probes), GLMMs used binomial distributions and a random intercept for each participant. Mind wandering GLMMs also adjusted for variable numbers of thought probes per participant. ANOVA chi-squared tests assessed main effects and interactions for GLMMs. The emmeans package (version 1.8.6) was used for post hoc z-tests, which were performed on the log-odds scale but, for clarity, back-transformed and reported as odds ratios (for which a value of 1 indicates no effect). Risk perception scores, which ranged continuously from 0 to 1, were analyzed with ANOVA F-tests and post hoc t-tests.

We also examined the effect of platform on multitasking. A chi-squared test and post hoc z-tests compared proportions of media-related multitasking, media-unrelated multitasking, and no multitasking responses between platforms. Low multitasking in the Prolific sample prohibited comparisons between conditions and analyses of its subjective impact on attention and accuracy.

Platform differences in individual traits were also examined using MANOVA F-tests and post hoc t-tests. Furthermore, we examined correlations between age, individual traits, accuracy, mind wandering, and risk perception within each platform. Spearman's rank-order correlations were used since dependent variables differed in scale. We also collapsed across condition because its effects were marginal in prior analyzes.

To examine the impact of our methodological adjustments in E2, we compared effect sizes of platform on accuracy, mind wandering, and multitasking between studies with interaction contrasts. We also compared the effect sizes of condition on risk perception between studies. Sensitivity analyses verified the consistency of results after excluding participants with accuracy scores of zero and controlling for significant differences in demographics between platforms. Results remained the same unless stated otherwise.

#### Results

#### Demographics

Supplementary Table S1 shows demographic characteristics for the two platform samples in E1. Comparisons revealed significant differences in education and employment. Compared to the Prolific sample, the MTurk sample was more educated and likely to be working full-time. Supplementary Table S2 shows demographic characteristics of the two platform samples in E2. Comparisons revealed significant differences in age and employment, but not education. The MTurk sample was found to be older and more educated than the Prolific sample.

Figure 2 displays the flow of participants through E1 and E2. Chi-squared test results for dropout rates revealed a non-significant difference between platforms in E1,  $\chi^2(1) = 0.00$ , p = 0.986, but significantly more dropouts for Prolific in E2,  $\chi^2(1) = 7.30$ , p = 0.006. In E1, average experiment completion time (from accepting to submitting the HIT or assignment) was 44.4 (SD = 16.9) min for Prolific and 69.9 (SD = 22.8) min for MTurk. In E2, average completion time was 44.2 (SD = 17.2) min for Prolific and 67.6 (SD = 22.4) min for MTurk.

#### Experiment 1



Experiment 2





Completion times were significantly longer for MTurk in both E1, *t*(140) = 7.96, *p* < 0.001, *d* = 1.28, 95% CI [0.93, 1.61], and E2, *t*(99.2) = 6.56, *d* = 1.20, 95% CI [0.83, 1.58].

#### Differences in accuracy by condition and platform

Figure 3 displays raw means and confidence intervals for accuracy, mind wandering, and risk perception by condition, platform, and experiment. In E1, ANOVA results for accuracy revealed a non-significant main effect of condition,  $\chi^2(1) = 2.48$ , p = 0.115, a significant main effect of platform,  $\chi^2(1) = 80.1$ , p < 0.001, and a significant platform by condition interaction,  $\chi^2(1) = 4.50$ , p = 0.034. Post hoc z-tests revealed significantly lower accuracy in the MTurk sample compared to the Prolific sample overall, OR = 0.05, 95% CI [0.03, 0.08]. Accuracy was also found to be significantly higher in the high-risk condition compared to the low-risk condition in the MTurk sample, OR = 3.10, 95% CI [1.30, 7.38], but not in the Prolific sample, OR = 0.84, 95% CI [0.37, 1.91]. In E2, ANOVA results for accuracy revealed a non-significant main effect of condition,  $\chi^2(1) = 0.21$ , p = 0.651, a significant main effect of platform,  $\chi^2(1) = 11.4$ , p = 0.001, and a non-significant condition by platform interaction,  $\chi^2(1) = 0.88$ , p = 0.348. A post hoc z-test revealed significantly lower accuracy for MTurk compared to Prolific overall, OR = 0.38, 95% CI [0.22, 0.66]. Comparing the effect of platform between experiments revealed





a significant decrease in accuracy differences between platforms from E1 to E2, z = 4.80, p < 0.001, OR = 7.36, 95% CI [3.26, 16.6].

#### Differences in mind wandering by condition and platform

In E1, ANOVA results for mind wandering revealed a non-significant main effect of condition,  $\chi^2(1) = 0.20$ , p = 0.655, a significant main effect of platform,  $\chi^2(1) = 13.9$ , p < 0.001, and a non-significant condition by platform interaction,  $\chi^2(1) = 1.99$ , p = 0.158. A post hoc z-test revealed significantly more mind wandering for MTurk compared to Prolific overall, OR = 4.36, 95% CI [2.03, 9.37]. In E2, ANOVA results for mind wandering revealed a non-significant main effect of condition,  $\chi^2(1) = 0.002$ , p = 0.968, a non-significant main effect of platform,  $\chi^2(1) = 0.67$ , p = 0.413, and a non-significant condition by platform interaction,  $\chi^2(1) = 0.34$ , p = 0.558. Comparing the effect of platform between experiments revealed a trending decrease in mind wandering differences between platforms from E1 to E2, z = -1.95, p = 0.051, OR = 0.33, 95% CI [0.11, 1.01]. However, this effect weakened after controlling for accuracy scores of zero, z = -0.91, p = 0.360, OR = 0.58, 95% CI [0.18, 1.86].

#### Differences in risk perception by condition and platform

In E1, ANOVA results for risk perception revealed a non-significant main effect of condition, F(1, 156) = 2.38, p = 0.125, a non-significant main effect of platform, F(1, 156) = 0.01, p = 0.916, and a non-significant condition by platform interaction, F(1, 156) = 0.43, p = 0.512. In E2, ANOVA results for risk perception revealed a significant main effect of condition, F(1, 129) = 24.5, p < 0.001, a non-significant main effect of platform, F(1, 129) = 0.02, p = 0.888, and a non-significant condition by platform interaction, F(1, 129) = 1.21, p = 0.273. A post hoc t-test revealed significantly higher perceived risk for the high-risk condition compared to the low-risk condition across platforms, d = 0.94, 95% CI [0.56, 1.31]. Comparing the effect of condition between experiments revealed

a significant increase in risk perception differences between conditions from E1 to E2, t(289) = 3.35, p = 0.001, d = 0.08, 95% CI [0.03, 0.13].

#### Differences in multitasking between platforms

Figure 4 displays raw proportions of multitasking responses by platform and experiment. It also displays Likert responses to questions about the perceived impact of multitasking on attention and accuracy in the task. In E1, chi-squared results revealed a significant difference in multitasking between platforms,  $\chi^2 = 35.0$ , p < 0.001. Post hoc z-tests revealed significantly more media-related multitasking for MTurk, OR = 19.8, 95% CI [5.73, 68.1], but no significant difference in multitasking between platforms,  $\chi^2 = 8.99$ , p = 0.004. A post hoc z-test revealed a significant difference in multitasking for MTurk, OR = 5.32, 95% CI [1.63, 17.3]. There were no reports of media-unrelated multitasking across samples. Comparing the effect of platform between experiments revealed a non-significant change in media-related multitasking differences between platforms from E1 to E2, z = -0.82, p = 0.412, OR = 0.53, 95% CI [0.12, 2.41].



**Figure 4.** Self-reported rate and impact of multitasking by recruitment platform and experiment. Multitasking was assessed by asking participants, "While completing this study, were you engaged in any media-related activities outside of the contents of the experiment?" Response options included: (1) "Yes"; (2) "No, I was not engaged in any activities outside of the contents of this study"; and (3) "No, but I was engaged in other, media-unrelated activities while completing this study." If participants selected options 1 (media-related multitasking) or 3 (media-unrelated multitasking), they were asked to what extent the other activities: (a) consumed attention; and (b) affected performance. Participants were assured that they would not be penalized for responding one way or another.

#### Differences in individual traits between platforms

In E1, MANOVA results for individual differences revealed a significant effect of platform, V = 0.46, F(4, 158) = 33.3, p < 0.001. Results from post hoc t-tests, displayed in Supplementary Table S3, revealed significantly higher scores on the ARCES, MAAS, BSSS, and WCFS for MTurk compared to Prolific. In E2, MANOVA results for individual differences revealed a significant effect of platform, V = 0.17, F(4, 131) = 6.71, p < 0.001. Results from post hoc t-tests, displayed in Supplementary Table S3, revealed significantly higher scores for MTurk compared to Prolific on the BSSS and WCFS, but not the ARCES or MAAS. Comparing the effect of platform between experiments revealed a significant decrease in individual trait differences between platforms from E1 to E2, t(289) = -4.93, p < 0.001, d = -1.16, 95% CI [-1.63, -0.70].

#### Correlations between individual traits, task performance, and mind wandering

Figure 5 shows correlations between age, individual traits, accuracy, mind wandering, and risk perception by platform and experiment. Upper triangles reflect correlations for Prolific samples while lower triangles reflect correlations for MTurk samples. Correlation coefficients for E1 and E2 can be found in Supplementary Tables S4 and S5, respectively. In both E1 platform samples, trait inattention (MAAS) correlated with cognitive errors in everyday life (ARCES) and work (WCFS) as well as mind wandering in the task (MW). Furthermore, in both samples, mind wandering negatively correlated with accuracy (Acc) and positively correlated with risk perception (Risk). More correlations were significant and generally stronger for MTurk compared to Prolific. In the MTurk sample, risk tolerance (BSSS) positively correlated with trait inattention, cognitive errors in everyday life and work, as well as mind wandering. Furthermore, only for MTurk were traits (ARCES, WCFS, MAAS, and BSSS) negatively correlated with accuracy. After removing participants with accuracy scores of zero, the MTurk sample lost a correlation between risk perception and mind wandering.

In E2, common to both samples were positive correlations between cognitive errors in everyday life (ARCES) and work (WCFS) as well as trait inattention (MAAS). Furthermore, in both samples, trait inattention positively correlated with mind wandering in the task (MW) and negatively correlated with accuracy (Acc). For Prolific, mind wandering positively correlated with cognitive errors in life and work, but negatively correlated with accuracy. For MTurk, accuracy negatively correlated with cognitive errors in life and work as well as risk tolerance (BSSS). Furthermore, for MTurk, risk tolerance positively correlated with trait inattention as well as cognitive errors in life and work. After controlling for accuracy scores of zero, for Prolific, mind wandering lost a negative correlation with cognitive errors at work but gained one with cognitive errors in life.



**Figure 5.** Correlations between individual traits, task performance, mind wandering, and risk perception for Prolific (upper triangles) and MTurk (lower triangles) by experiment. Acc=accuracy (percentage of correct responses to targets), ARCES = Attention-Related Cognitive Errors Scale, BSSS = Brief Sensation Seeking Scale, MAAS = Mindful Awareness of Attention Scale, MW = mind wandering (percentage of off-task reports to thought probes), WCFS = Workplace Cognitive Failures Scale, Risk = risk perception (perception of the risk of losing points for task errors). Results are collapsed across low-risk and high-risk conditions. Coefficients were generated with Spearman's rank-order method. \*p < .05, \*\*p < .01, \*\*\*p < .001.

Scientific Reports | (2023) 13:20574 |

#### Discussion

The most striking findings of the present study relate to its second aim, which was to compare disengagement in an attention task between Prolific and MTurk. Greater mind wandering, media-related multitasking, and lower task performance provide convergent evidence for greater disengagement on MTurk compared to Prolific. These findings corroborate observations from survey studies showing significantly fewer correct responses to attentioncheck questions in MTurk samples<sup>38,41,64</sup>. A novel contribution of the present study, however, is that it examined disengagement in terms of attention-task performance as well as self-reports of mind wandering and multitasking. This is relevant since online recruitment platforms are now commonly used for studies involving cognitive tasks<sup>48-54</sup>. Furthermore, most previous comparisons of these platforms were based on survey data, which may be subject to factors, such as non-naivete<sup>40,43</sup>, that are less applicable in the context of cognitive tasks. It is notable, however, that the effects of platform significantly decreased, with differences in mind wandering losing significance, when we modified MTurk recruitment methods in E2. This suggests that offering higher remuneration may have motivated MTurk participants to stay engaged and/or increasing reputation criteria yielded a more attentive sample. Nevertheless, in contrast to the finding of Rioja and colleagues<sup>55</sup>, there remained a significant difference in task performance between platforms. Our results may have differed from theirs because the task we used was longer ( $\approx 25$  min vs.  $\approx 10$  min) and more difficult (working memory vs. vigilance), thus potentially more sensitive to attentional disengagement. Psychology researchers may be drawn to Prolific, or services like CloudResearch that implement similar controls on MTurk<sup>41,42,47</sup>, for studies involving cognitive tasks, since full engagement is often assumed in these contexts<sup>65,66</sup>. However, disengagement in MTurk samples may be more reflective of everyday life<sup>67,68</sup>, which could be of value for ecological validity and generalizability.

In addition to between-platform differences in task disengagement during the study, we also observed differences in individual traits related to disengagement and risk tolerance. These findings align with results from survey studies. For instance, Arechar and Rand<sup>69</sup> presented data suggesting that MTurk participants tend to be more dishonest than Prolific participants. Like in previous studies, we also found significant differences in demographic characteristics, namely education and employment in E1 and age and employment in E2, between the MTurk and Prolific samples<sup>40,70,71</sup>. Given that scores on all of the individual trait questionnaires were higher for MTurk compared to Prolific in E1, these differences could reflect response biases more than actual variability in individual traits between platforms. At the same time, between-platform differences in individual traits related to disengagement decreased along with task-related indicators of disengagement (mind wandering and accuracy) from E1 to E2, suggesting that increasing remuneration, or more likely reputation criteria, may have selected against MTurk participants high in traits related to disengagement.

Across E1 and E2, we found differences in correlations among individual traits, task performance, mind wandering, and risk perception between the two platforms. While the direction of relations was consistent, correlations were generally stronger and more of them significant for MTurk compared to Prolific. A similar observation was made in a study examining relations between trait impulsivity and gambling-like spending on loot boxes in video games<sup>72</sup>. The authors attributed these differences to higher data quality in the Prolific sample. Particularly notable in the present study was the relative lack of significant correlations between task accuracy and individual traits for Prolific compared to MTurk across E1 and E2. While such differences could arise from between-platform variation in the ranges of underlying data, our use of Spearman's rank-order method makes this an unlikely explanation. Alternatively, this divergence in relations between self-report and objective measures may reflect response bias. For instance, Prolific participants may be more compelled to present themselves as attentive, resulting in less concordance between their self-reported traits related to disengagement and objective task performance compared to MTurk participants. At the same time, mind wandering significantly correlated with individual traits and task accuracy for MTurk in E1, but not in E2, while these relations remained consistent for Prolific. Thus, higher remuneration and/or reputation criteria for MTurk may have resulted in a sample prone to reporting less disengagement in the attention task than suggested by their objective performance.

The first aim of the present study was to examine the effect of risk on attentional disengagement. Findings were mixed and relatively weak, however. We found partial support for the hypothesis that greater risk of negative consequences for poor task performance leads to lower attentional disengagement, but only in the MTurk sample of E1. Within this sample, participants in the high-risk condition exhibited higher task performance compared those in the low-risk condition. Contrary to our hypothesis, they did not report significantly less mind wandering or greater risk perception, however. The observed effect of risk on disengagement in the present study aligns with findings from traffic safety studies indicating that drivers are less likely to disengage attention from the road when the perceived risk of a crash is high<sup>30,31,73</sup>. Our study extends these correlational findings by demonstrating a causal relationship between risk and attentional disengagement. At the same time, in E2, we found a significant main effect of risk condition on risk perception across Prolific and MTurk samples, but not on task performance or mind wandering in either. Thus, results concerning the effects of risk on disengagement are inconclusive but may speak to the impact of recruitment platform and recruitment methods, in terms of remuneration and reputation criteria, on experimental findings from cognitive tasks. In this way, our findings align with those from survey-based studies revealing results that differ as a function of recruitment platform<sup>38,70,72</sup> and data quality controls (e.g., approval rating, HITs completed, attention checks)<sup>45,74,75</sup>.

A possible explanation for the persistence of differences between platforms relates to their differing participant compensation policies. Prolific explicitly outlines a policy regarding acceptable methods for researchers to ensure that they only pay for high-quality data. Specifically, Prolific permits the use of attention-check and comprehension-check questions to assess participant disengagement as a basis to deny compensation<sup>76</sup>. In contrast, MTurk policies do not discuss disengagement as a basis to deny compensation. It stands to reason that the presence of this policy on Prolific, and its absence on MTurk, could impact results, particularly for studies in which attention is a dependent variable. In alignment with our original hypothesis, Prolific's policy may reduce attentional disengagement by increasing the perceived risk of losing compensation for it. Supporting this notion, previous studies have found that offering incentives can motivate participants to decrease mind wandering in attention tasks<sup>77</sup>. Alternatively, Prolific's policy may select against participants with a higher trait-level propensity for disengagement. This could explain the between-platform differences in individual trait questionnaire scores, self-reported disengagement, and attention-task performance. At the same time, Prolific's policy may function as a demand characteristic that encourages a conservative response bias on individual trait questionnaires of disengagement, mind wandering probes, and questions about multitasking. This could explain why the Prolific sample had weak, non-significant correlations between traits related to disengagement and task performance, in contrast to the strong, significant correlations found between these variables in the MTurk sample. Further research is needed to explore these speculations.

Overall, increasing MTurk remuneration and reputation criteria from E1 to E2 reduced the size of differences between platforms across multiple study variables, but some new differences emerged. While dropout rates did not differ between Prolific and MTurk in E1, there were significantly fewer dropouts from MTurk in E2. This may relate to the increase in remuneration for the MTurk sample, since participants were only remunerated if they fully completed the experiment. Furthermore, while equivalent in terms of purchasing power parity, the remuneration offered to Prolific participants was relatively low for the platform, whereas that offered to MTurk participants was relatively high, which may have disproportionately motivated them to finish the experiment. We also observed a large number of E2 MTurk participants reporting an ineligible age in the demographics questionnaire. The increase in remuneration criteria may also have yielded a sample that was more adept at circumventing screening measures.

#### Conclusion

In the present study, we hypothesized that increased risk of negative consequences for poor task performance could decrease attentional disengagement. While investigating this hypothesis, we stumbled across surprising differences in participant disengagement, in terms of task performance, mind wandering, and individual traits between two popular online recruitment platforms, namely Prolific and MTurk. Across two experiments (E1 and E2), MTurk participants exhibited higher attentional disengagement in an attention task and distinct patterns of relations between individual trait and task variables compared to Prolific participants. Results of our risk manipulation also varied as a function of platform and MTurk recruitment method, which changed from E1 to E2. Increasing MTurk remuneration and reputation controls reduced, but did not eliminate, differences between platforms. Additional controls implemented on Prolific may explain this gap, potentially including their participant remuneration policy. Future research is needed to test this possibility. Nevertheless, researchers should be aware of how choices of recruitment platform and methods may impact results in the context of cognitive tasks.

#### Data availability

The datasets generated and/or analysed during the current study are available in the Open Science Framework repository, https://osf.io/k9a6c/.

Received: 17 April 2023; Accepted: 26 October 2023 Published online: 23 November 2023

#### References

- 1. Ralph, B. C. W., Wammes, J. D., Barr, N. & Smilek, D. Wandering minds and wavering goals: Examining the relation between mind wandering and grit in everyday life and the classroom. *Can. J. Exp. Psychol. Rev. Can. Psychol. Exp.* **71**, 120–132 (2017).
- Robison, M. K., Gath, K. I. & Unsworth, N. The neurotic wandering mind: An individual differences investigation of neuroticism, mind-wandering, and executive control. Q. J. Exp. Psychol. 70, 649–663 (2017).
- 3. Moran, C. N. et al. Young and restless, old and focused: Age-differences in mind-wandering frequency and phenomenology. Psychol. Aging 36, 252 (2021).
- 4. Seli, P. et al. Mind-wandering across the age gap: Age-related differences in mind-wandering are partially attributable to age-related differences in motivation. J. Gerontol. Ser. B 76, 1264–1271 (2021).
- Brosowsky, N. P., Murray, S., Schooler, J. W. & Seli, P. Thought dynamics under task demands: Evaluating the influence of task difficulty on unconstrained thought. J. Exp. Psychol. Hum. Percept. Perform. 47, 1298–1312 (2021).
- Seli, P., Konishi, M., Risko, E. F. & Smilek, D. The role of task difficulty in theoretical accounts of mind wandering. *Conscious. Cogn.* 65, 255–262 (2018).
- 7. Kane, M. J. et al. For whom the mind wanders, and when, varies across laboratory and daily-life settings. Psychol. Sci. 28, 1271–1289 (2017).
- Linz, R., Pauly, R., Smallwood, J. & Engert, V. Mind-wandering content differentially translates from lab to daily life and relates to subjective stress experience. *Psychol. Res.* 85, 649–659 (2021).
- 9. Drody, A. C., Pereira, E. J. & Smilek, D. A desire for distraction: Uncovering the rates of media multitasking during online research studies. *Sci. Rep.* 13, 781 (2023).
- Casler, K., Bickel, L. & Hackett, E. Separate but equal? A comparison of participants and data gathered via Amazon's MTurk, social media, and face-to-face behavioral testing. *Comput. Hum. Behav.* 29, 2156–2160 (2013).
- Clifford, S. & Jerit, J. Is there a cost to convenience? An experimental comparison of data quality in laboratory and online studies. J. Exp. Political Sci. 1, 120–131 (2014).
- 12. Crump, M. J. C., McDonnell, J. V. & Gureckis, T. M. Evaluating Amazon's Mechanical Turk as a tool for experimental behavioral research. *PLoS ONE* **8**, e57410 (2013).
- 13. Palan, S. & Schitter, C. Prolific. ac-A subject pool for online experiments. J. Behav. Exp. Finance 17, 22-27 (2018).
- 14. Carrier, L. M., Rosen, L. D., Cheever, N. A. & Lim, A. F. Causes, effects, and practicalities of everyday multitasking. *Dev. Rev.* 35, 64–78 (2015).
- Wammes, J. D. et al. Disengagement during lectures: Media multitasking and mind wandering in university classrooms. Comput. Educ. 132, 76–89 (2019).

- Kam, J. W. Y. & Handy, T. C. The neurocognitive consequences of the wandering mind: A mechanistic account of sensory-motor decoupling. Front. Psychol. 4, 725 (2013).
- Seli, P., Cheyne, J. A. & Smilek, D. Wandering minds and wavering rhythms: Linking mind wandering and behavioral variability. J. Exp. Psychol. Hum. Percept. Perform. 39, 1-5 (2013).
- Jana, S. & Aron, A. R. Mind wandering impedes response inhibition by affecting the triggering of the inhibitory process. *Psychol. Sci.* 33, 1068–1085 (2022).
- 19. Seli, P. The attention-lapse and motor decoupling accounts of SART performance are not mutually exclusive. *Conscious. Cogn.* 41, 189–198 (2016).
- Thomson, D. R., Seli, P., Besner, D. & Smilek, D. On the link between mind wandering and task performance over time. *Conscious. Cogn.* 27, 14–26 (2014).
- May, K. E. & Elder, A. D. Efficient, helpful, or distracting? A literature review of media multitasking in relation to academic performance. Int. J. Educ. Technol. High. Educ. 15, 13 (2018).
- Patterson, M. C. A naturalistic investigation of media multitasking while studying and the effects on exam performance. *Teach. Psychol.* 44, 51–57 (2017).
- Wallace, J. C. & Chen, G. Development and validation of a work-specific measure of cognitive failure: Implications for occupational safety. J. Occup. Organ. Psychol. 78, 615–632 (2005).
- 24. Galéra, C. et al. Mind wandering and driving: Responsibility case-control study. BMJ 345, e8105 (2012).
- 25. Baird, B. et al. Inspired by distraction: Mind wandering facilitates creative incubation. Psychol. Sci. 23, 1117–1122 (2012).
- Kapadia, C. & Melwani, S. More tasks, more ideas: The positive spillover effects of multitasking on subsequent creativity. J. Appl. Psychol. 106, 542–559 (2021).
- Baird, B., Smallwood, J. & Schooler, J. W. Back to the future: Autobiographical planning and the functionality of mind-wandering. Conscious. Cogn. 20, 1604–1611 (2011).
- Kvavilashvili, L. & Rummel, J. On the nature of everyday prospection: A review and theoretical integration of research on mindwandering, future thinking, and prospective memory. *Rev. Gen. Psychol.* 24, 210–237 (2020).
- Ralph, B. C. W., Smith, A. C., Seli, P. & Smilek, D. The relation between task-unrelated media multitasking and task-related motivation. *Psychol. Res.* 85, 408–422 (2021).
- Harbeck, E. L. & Glendon, A. I. Driver prototypes and behavioral willingness: Young driver risk perception and reported engagement in risky driving. J. Saf. Res. 66, 195–204 (2018).
- 31. Zhou, R., Zhang, Y. & Shi, Y. Driver's distracted behavior: The contribution of compensatory beliefs increases with higher perceived risk. *Int. J. Ind. Ergon.* 80, 103009 (2020).
- 32. Carter, P. M., Bingham, C. R., Zakrajsek, J. S., Shope, J. T. & Sayer, T. B. Social norms and risk perception: Predictors of distracted driving behavior among novice adolescent drivers. J. Adolesc. Health 54, S32–S41 (2014).
- Cordellieri, P. et al. Young drivers 'on the phone: Personality traits, driving attitudes, beliefs, and risk perception on using mobile phone while driving. Psychol. Stud. 67, 459–467 (2022).
- 34. Pope, C. N., Bell, T. R. & Stavrinos, D. Mechanisms behind distracted driving behavior: The role of age and executive function in the engagement of distracted driving. Accid. Anal. Prev. 98, 123–129 (2017).
- Starkey, N. J. & Isler, R. B. The role of executive function, personality and attitudes to risks in explaining self-reported driving behaviour in adolescent and adult male drivers. *Transp. Res. Part F Traffic Psychol. Behav.* 38, 127–136 (2016).
- Burdett, B. R. D., Charlton, S. G. & Starkey, N. J. Mind wandering during everyday driving: An on-road study. Accid. Anal. Prev. 122, 76–84 (2019).
- O'Neill, K., Smith, A. P., Smilek, D. & Seli, P. Dissociating the freely-moving thought dimension of mind-wandering from the intentionality and task-unrelated thought dimensions. *Psychol. Res.* 85, 2599–2609 (2021).
- Adams, T. L., Li, Y. & Liu, H. A replication of beyond the Turk: Alternative platforms for crowdsourcing behavioral research— Sometimes preferable to student groups. AIS Trans. Replication Res. 6, 15 (2020).
- Chandler, J., Rosenzweig, C., Moss, A. J., Robinson, J. & Litman, L. Online panels in social science research: Expanding sampling methods beyond Mechanical Turk. *Behav. Res. Methods* 51, 2022–2038 (2019).
- 40. Peer, E., Brandimarte, L., Samat, S. & Acquisti, A. Beyond the Turk: Alternative platforms for crowdsourcing behavioral research. *J. Exp. Soc. Psychol.* **70**, 153–163 (2017).
- Peer, E., Rothschild, D., Gordon, A., Evernden, Z. & Damer, E. Data quality of platforms and panels for online behavioral research. Behav. Res. Methods 54, 1643–1662 (2021).
- Litman, L., Moss, A., Rosenzweig, C. & Robinson, J. Reply to MTurk, Prolific or panels? Choosing the right audience for online research. SSRN Electron. J. https://doi.org/10.2139/ssrn.3775075 (2021).
- Chandler, J., Mueller, P. & Paolacci, G. Nonnaïveté among Amazon Mechanical Turk workers: Consequences and solutions for behavioral researchers. *Behav. Res. Methods* 46, 112–130 (2014).
- 44. Keith, M. G., Tay, L. & Harms, P. D. Systems perspective of Amazon Mechanical Turk for organizational research: Review and recommendations. *Front. Psychol.* **8**, 1359 (2017).
- Lu, L., Neale, N., Line, N. D. & Bonn, M. Improving data quality using Amazon Mechanical Turk through platform setup. Cornell Hosp. Q. 63, 231–246 (2022).
- Robinson, J., Rosenzweig, C., Moss, A. J. & Litman, L. Tapped out or barely tapped? Recommendations for how to harness the vast and largely unused potential of the Mechanical Turk participant pool. *PLoS ONE* 14, e0226394 (2019).
- 47. Zickar, M. J. & Keith, M. G. Innovations in sampling: Improving the appropriateness and quality of samples in organizational research. *Annu. Rev. Organ. Psychol. Organ. Behav.* **10**, 315–337 (2023).
- Calleja, M. O. & Willoughby, A. R. The effects of search-irrelevant working memory content on visual search. Atten. Percept. Psychophys. 85, 293–300 (2023).
- 49. Kruger, T. B., Marty-Dugas, J., Ralph, B. C. W., Dixon, M. J. & Smilek, D. The relation between self-reported healthy living and attentional engagement in everyday life. *Curr. Res. Behav. Sci.* 4, 100086 (2023).
- Kelm, M. R., Diercks, C. M., Dunning, E. D. & Lunkenheimer, E. Parental working memory buffers associations between COVID-19 hardships and child mental health. J. Appl. Dev. Psychol. 87, 101559 (2023).
- 51. Drody, A. C., Ralph, B. C. W., Danckert, J. & Smilek, D. Boredom and media multitasking. Front. Psychol. 13, 1214 (2022).
- Chua, K. Q. et al. Relationship between contentment and working memory capacity: Experimental and naturalistic evidence. Curr. Psychol. https://doi.org/10.1007/s12144-022-03714-7 (2022).
- 53. Smith, A. P. et al. Fixation, flexibility, and creativity: The dynamics of mind wandering. J. Exp. Psychol. Hum. Percept. Perform. 48, 689–710 (2022).
- 54. Wessel, J. R., Jiang, J. & Stolley, J. J. Action errors impair active working memory maintenance. J. Exp. Psychol. Gen. 151, 1325–1340 (2022).
- Rioja, K., Cekic, S., Bavelier, D. & Baumgartner, S. E. Unraveling the link between media multitasking and attention across three samples. *Technol. Mind Behav.* 4, tmb0000106 (2023).
- 56. Kennedy, R. et al. The shape of and solutions to the MTurk quality crisis. Political Sci. Res. Methods 8, 614-629 (2020).
- Peer, E., Vosgerau, J. & Acquisti, A. Reputation as a sufficient condition for data quality on Amazon Mechanical Turk. *Behav. Res. Methods* 46, 1023–1031 (2014).

- 58. Vo, H. L. & Vo, D. H. The purchasing power parity and exchange-rate economics half a century on. J. Econ. Surv. 37, 446-479 (2023).
- Cheyne, J. A., Carriere, J. S. A. & Smilek, D. Absent-mindedness: Lapses of conscious awareness and everyday cognitive failures. Conscious. Cogn. 15, 578–592 (2006).
- 60. Brown, K. W. & Ryan, R. M. The Mindful Attention Awareness Scale (MAAS). (2003).
- Carriere, J. S. A., Cheyne, J. A. & Smilek, D. Everyday attention lapses and memory failures: The affective consequences of mindlessness. *Conscious. Cogn.* 17, 835–847 (2008).
- 62. Hoyle, R. H., Stephenson, M. T., Palmgreen, P., Lorch, E. P. & Donohew, R. L. Reliability and validity of a brief measure of sensation seeking. *Personal. Individ. Differ.* 32, 401–414 (2002).
- 63. Konishi, M., McLaren, D. G., Engen, H. & Smallwood, J. Shaped by the past: The default mode network supports cognition that is independent of immediate perceptual input. *PLoS ONE* **10**, e0132209 (2015).
- Douglas, B. D., Ewell, P. J. & Brauer, M. Data quality in online human-subjects research: Comparisons between MTurk, Prolific, CloudResearch, Qualtrics, and SONA. *PLoS ONE* 18, e0279720 (2023).
- 65. Medina, J. M., Wong, W., Díaz, J. A. & Colonius, H. Advances in modern mental chronometry. Front. Hum. Neurosci. 9, 256 (2015).
- 66. Posner, M. I. Chronometric Explorations of Mind xiii–271 (Lawrence Erlbaum, 1978).
- 67. Killingsworth, M. A. & Gilbert, D. T. A wandering mind is an unhappy mind. Science 330, 932-932 (2010).
- 68. Seli, P. et al. How pervasive is mind wandering, really?. Conscious. Cogn. 66, 74-78 (2018).
- 69. Arechar, A. A. & Rand, D. G. Turking in the time of COVID. Behav. Res. Methods 53, 2591-2595 (2021).
- Matsuura, T., Hasegawa, A. A., Akiyama, M. & Mori, T. Careless participants are essential for our phishing study: Understanding the impact of screening methods. In *European Symposium on Usable Security 2021* 36–47 (ACM, 2021). https://doi.org/10.1145/ 3481357.3481515.
- Oppenlaender, J., Milland, K., Visuri, A., Ipeirotis, P. & Hosio, S. Creativity on paid crowdsourcing platforms. In 1–14 (ACM, 2020). https://doi.org/10.1145/3313831.3376677.
- 72. Garrett, E. P. et al. Impulsivity and loot box engagement. Telemat. Inform. 78, 101952 (2023).
- 73. Nguyen-Phuoc, D. Q., Oviedo-Trespalacios, O., Su, D. N., De Gruyter, C. & Nguyen, T. Mobile phone use among car drivers and motorcycle riders: The effect of problematic mobile phone use, attitudes, beliefs and perceived risk. *Accid. Anal. Prev.* **143**, 105592 (2020).
- 74. Chmielewski, M. & Kucker, S. C. An MTurk crisis? Shifts in data quality and the impact on study results. Soc. Psychol. Personal. Sci. 11, 464–473 (2020).
- 75. Pyo, J. & Maxfield, M. G. Cognitive effects of inattentive responding in an MTurk sample. Soc. Sci. Q. 102, 2020–2039 (2021).
- Prolific's Attention and Comprehension Check Policy. *Prolific* https://researcher-help.prolific.co/hc/en-gb/articles/3600092235 53-Prolific-s-Attention-and-Comprehension-Check-Policy (2023).
- Seli, P., Schacter, D. L., Risko, E. F. & Smilek, D. Increasing participant motivation reduces rates of intentional and unintentional mind wandering. *Psychol. Res.* 83, 1057–1069 (2019).

#### Acknowledgements

This study was supported by the Natural Sciences and Engineering Research Council of Canada (RGPIN-2019-04071) and a Mitacs Accelerate grant with Exo Insights (IT25494). The lead author was supported by a postdoctoral fellowship from the Lupina Foundation. None of these funders were involved in: designing the study; collecting, analyzing, or interpreting the data; writing the report; or the decision to submit the article for publication.

#### Author contributions

D.A.A. contributed to conceptualization, methodology, investigation, data curation, formal analysis, writing, review and editing, and visualization. D.S. contributed to conceptualization, methodology, review and editing, supervision, funding acquisition, and resources.

#### **Competing interests**

The authors declare no competing interests.

#### Additional information

**Supplementary Information** The online version contains supplementary material available at https://doi.org/10.1038/s41598-023-46048-5.

Correspondence and requests for materials should be addressed to D.A.A.

Reprints and permissions information is available at www.nature.com/reprints.

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2023