Supplementary Materials: Effects of human-animal interactions on affect and cognition

Elise R. Thayer[1] & Jeffrey R. Stevens[1]

[1] University of Nebraska-Lincoln

Supplementary Methods

Deese-Roedinger-McDermott long-term memory

In the presentation phase of the Deese-Roedinger-McDermott long-term memory task, we presented two sets of ten words associated with a not-presented *critical word* (e.g., child). For the recall phase, participants were told that they would view some of the words they were instructed to remember at the beginning of the experiment. They viewed 6 of the 20 previously presented words, the two critical words, and four unrelated words in random order (Table S2) and responded via key press whether they recalled seeing the words previously in the experiment. We calculated d' as $z(P(recall|presented\ word)) - z(P(recall|not\ presented\ word))$. A value of 0 indicates that a participant was not able to discriminate between previously presented and not presented words and larger values indicate greater propensities to discriminate between presented and not-presented words (Haatveit et al., 2010).

Backwards digit span

Following a practice sequence of numbers, participants responded to a total of 14 unique sequences that increased in span (number of digits), starting with three digits and ending with nine digits. We pseudo-randomly generated each sequence such that no digit was repeated within a given sequence.

N-back

Each letter was presented for 0.5 seconds and followed by a 2.5 second inter-stimulus interval (fixation cross). Participants completed a brief practice before viewing a total of 96 letters, of which 16-19 were targets (i.e., matching the stimulus from two trials prior). We calculated d' as $z(P(hit|target)) - z(P(hit|not\ target))$ (Haatveit et al., 2010).

Cardiovascular activity

Participants wore an Empatica E4 wristband on their left wrist for the duration of the experiment to provide a continuous measure of cardiovascular activity, specifically heart rate variability (McCarthy et al., 2016; Empatica, 2018). The E4 wristband uses a photoplethysmogram (PPG) sensor located on the backside of the device face seated atop the wearer's wrist that samples heart beats at 64 Hertz. We do not provide analyses of cardiovascular data because the device did not sample at a sufficient frequency to detect fine changes in heart rate variability (Shaffer & Ginsberg, 2017).

HAI condition

The dog used for the HAI condition was a 65-pound, male Catahoula leopard mix that was Canine Good Citizen certified. When not interacting with participants, the dog was housed in JRS's office or adjacent experiment room that were both equipped with a large kennel and access to water. The dog participated in 1-5 sessions spread throughout the day with at least 40 minutes between sessions.

Participants remained seated for the duration of the interaction. If participants asked questions of or made conversation with the researcher, the researcher briefly informed participants that they would be able to discuss more at the end of the experiment. The dog remained in the experiment room within visual contact after the animal interaction period to rest next to the researcher.

After the experiment ended, researchers debriefed participants who experienced the HAI condition. The researcher explained that participants were not made aware of the dog in recruitment or consent materials to avoid biasing the sample and the results. Researchers requested that participants not discuss the presence of the dog in the experiment with other individuals to help maintain an unbiased sample. Then, researchers invited participants to ask any questions about the nature of the study.

Analysis

We used R (Version 4.0.2; R Core Team, 2020) and the R-packages *BayesFactor* (Version 0.9.12.4.2; Morey & Rouder, 2018), *ggbeeswarm* (Version 0.6.0; Clarke & Sherrill-Mix, 2017), *ggcorrplot* (Version 0.1.3; Kassambara, 2019), *here* (Version 0.1; Müller, 2017), *Hmisc* (Version 4.4.0; Harrell Jr et al., 2020), *lsmeans* (Version 2.30.0; Lenth, 2016), *papaja* (Version 0.1.0.9997; Aust & Barth, 2020), *patchwork* (Version 1.0.1; Pedersen, 2020), *psych* (Version 1.9.12.31; Revelle, 2019), *rcompanion* (Version 2.3.25; Mangiafico, 2020), and *tidyverse* (Version 1.3.0; Wickham et al., 2019) for all our analyses. We prepared the manuscript using *rmarkdown* (Version 2.3; Xie et al., 2018).

To assess potential condition differences pre-condition, we conducted independent samples t-tests or Wilcoxon rank sum tests in the case of test violations. For condition effects, we used analysis of covariance to examine the effects of condition on post-scores controlling for pre-scores. We report predicted marginal means and 95% confidence intervals and calculated effect sizes with generalized eta squared (η^2_G). We used both visual and test-based methods to check analysis of covariance test assumptions: linearity between covariate and outcome variable, homogeneity of regression slopes, normally distributed residuals on outcome variable, and homogeneity of variance (Johnson, 2016). Since we recorded responses only at post-condition for the long-term memory task, we compared between-groups differences with a Wilcoxon rank sum test, as it violated assumptions of an independent samples t-test. For these, we calculated effect sizes with r, the z-score of the test statistic divided by the total number of observations.

In Experiment 1, the analyses of covariance for negative affect and Necker cube significantly violated model assumptions, so we log-transformed negative affect pre- and post-scores and removed outliers (standard deviation $\times$ 3; 2 observations) from Necker post-scores to conform to model assumptions. We excluded one observation from DRM (participant expressed comprehension issues after task), three from Necker cube (participants expressed comprehension issues after task), and three from digit span (participants did not request response sheet) when running respective analyses, inserting missing data for these subjects. Similarly, in Experiment 2, we log-transformed negative affect pre- and post-scores and removed a single outlier (standard deviation $\times$ 3) to conform to analysis of covariance model assumptions. We excluded one observation from DRM (participant expressed comprehension issues after task), three from Necker cube (participants expressed comprehension issues after task), six from digit span (participants did not request response sheet or did not complete more than half of task), and three from n-back (two participants did not record any responses and one participant expressed comprehension issues after task).

We supplemented the frequentist analyses with Bayes factors and drew inferences based on Bayes factors, that is, the strength of evidence for the alternative over the null hypothesis. For example, $BF = 15$ suggests that there is 15 times more evidence for the alternative than the null hypothesis. Bayes factors between 3-10 provide moderate evidence for the alternative hypothesis, those between 10-30 provide strong evidence, those between 30-100 provide very strong evidence, and those above 100 provide extreme evidence; reciprocal values (1/3, 1/10, 1/30, 1/100) provide comparable evidence for the null hypothesis (Wagenmakers et al., 2018). We used the *BayesFactor* package to calculate most Bayes factors, except for the Wilcoxon rank sum tests which we calculated from code provided by van Doorn et al. (2020).

Table S1
Demographics

Measure	Experiment 1 N (%)	Experiment 2 N (%)
Gender		
Age (M(SD))	19.0 (1.4)	20.0 (1.8)
Female	60 (82.2)	66 (79.5)
Male	13 (17.8)	17 (20.5)
Ethnicity		
Other	0 (0)	0 (0)
Asian	6 (8.2)	9 (10.8)
Black	5 (6.8)	5 (6)
Native American	0 (0)	3 (3.6)
Pacific Islander	0 (0)	0 (0)
White/Caucasian	58 (79.5)	58 (69.9)
Family Income		
Other	4 (5.5)	8 (9.6)
< \$25000	3 (4.1)	5 (6)
\$25000-\$50000	6 (8.2)	12 (14.5)
\$50000-\$75000	9 (12.3)	15 (18.1)
\$75000-\$100000	17 (23.3)	13 (15.7)
> \$100000	32 (43.8)	31 (37.3)
Preferred not to answer	6 (8.2)	7 (8.4)

Table S2
Words used in Deese-Roedinger-McDermott long-term memory test.

Presentation words	Recall words
kid	kid*
adult	toy*
adolescent	immature*
toy	beaker*
parent	physics*
baby	test tube*
dependent	child
immature	chemistry
brat	blouse
juvenile	table
beaker	victory
element	cardboard
lab	
physics	
formula	
molecule	
flask	
test tube	
scientist	
electron	

* Denotes recall words that were present in presentation phase.

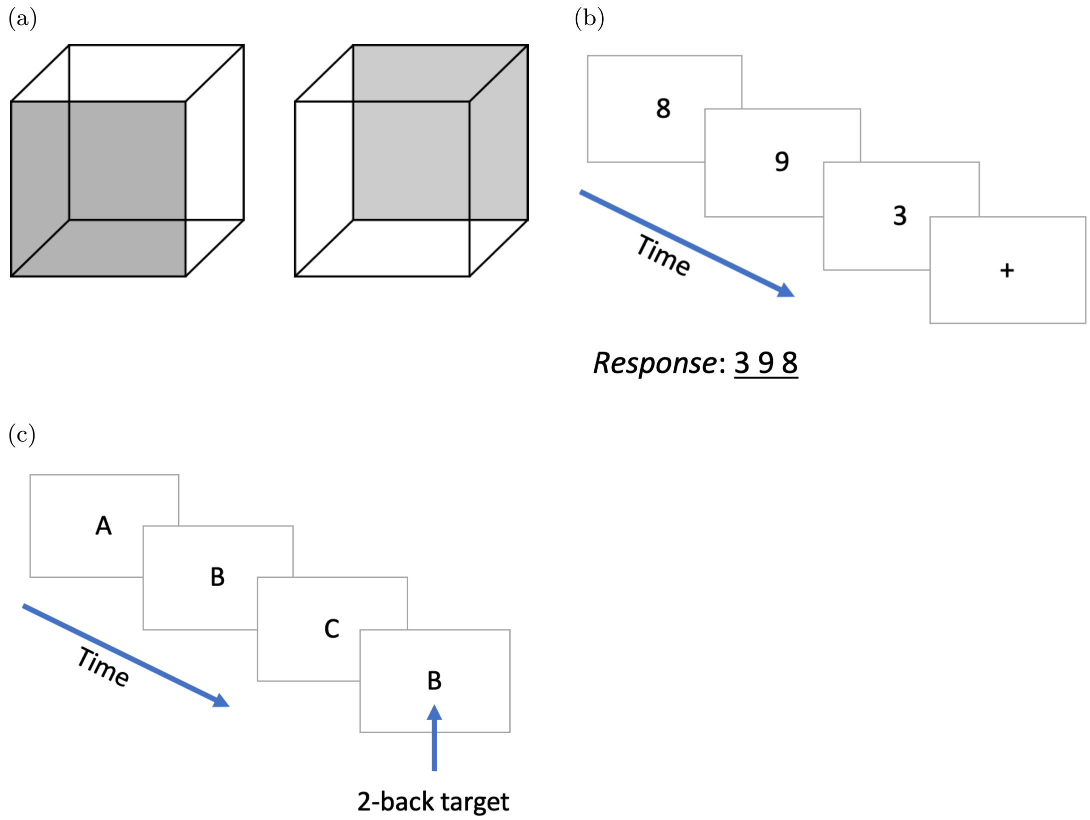

Figure S1. Cognitive tasks: (a) Necker Cube Pattern Control Test, (b) backwards digit span test, and (c) n-back task.

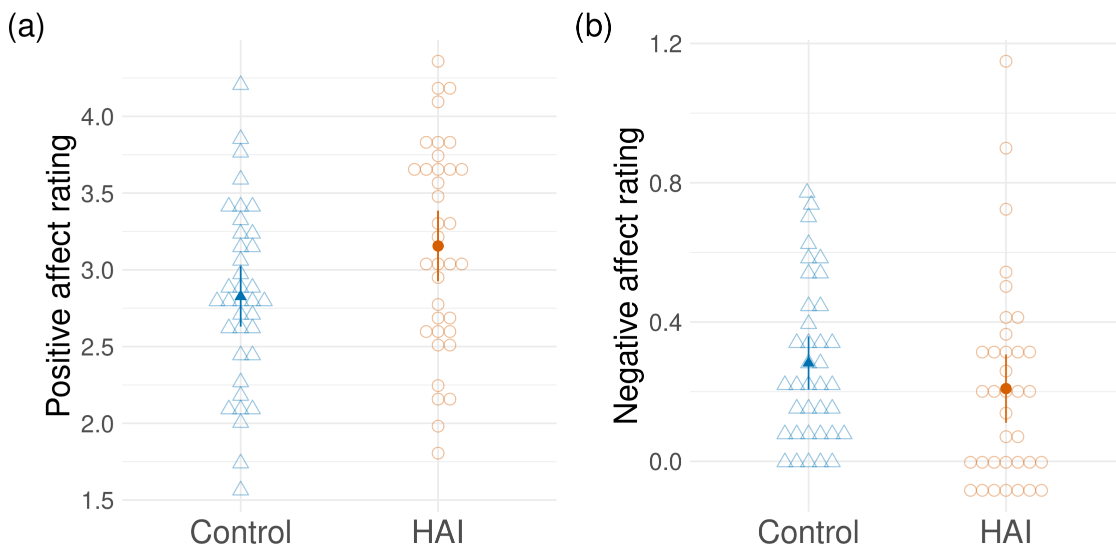

Figure S2. Post-condition predicted affect scores (controlling for pre-condition scores) for control and HAI (human-animal interaction) groups in Experiment 1. Scores show (a) positive PANAS ratings and (b) negative PANAS ratings. Negative affect scores are log-transformed. Open triangles represent individual control participant scores, open circles represent individual HAI participant scores, filled triangles and circles represent condition group means, error bars represent 95% confidence intervals.

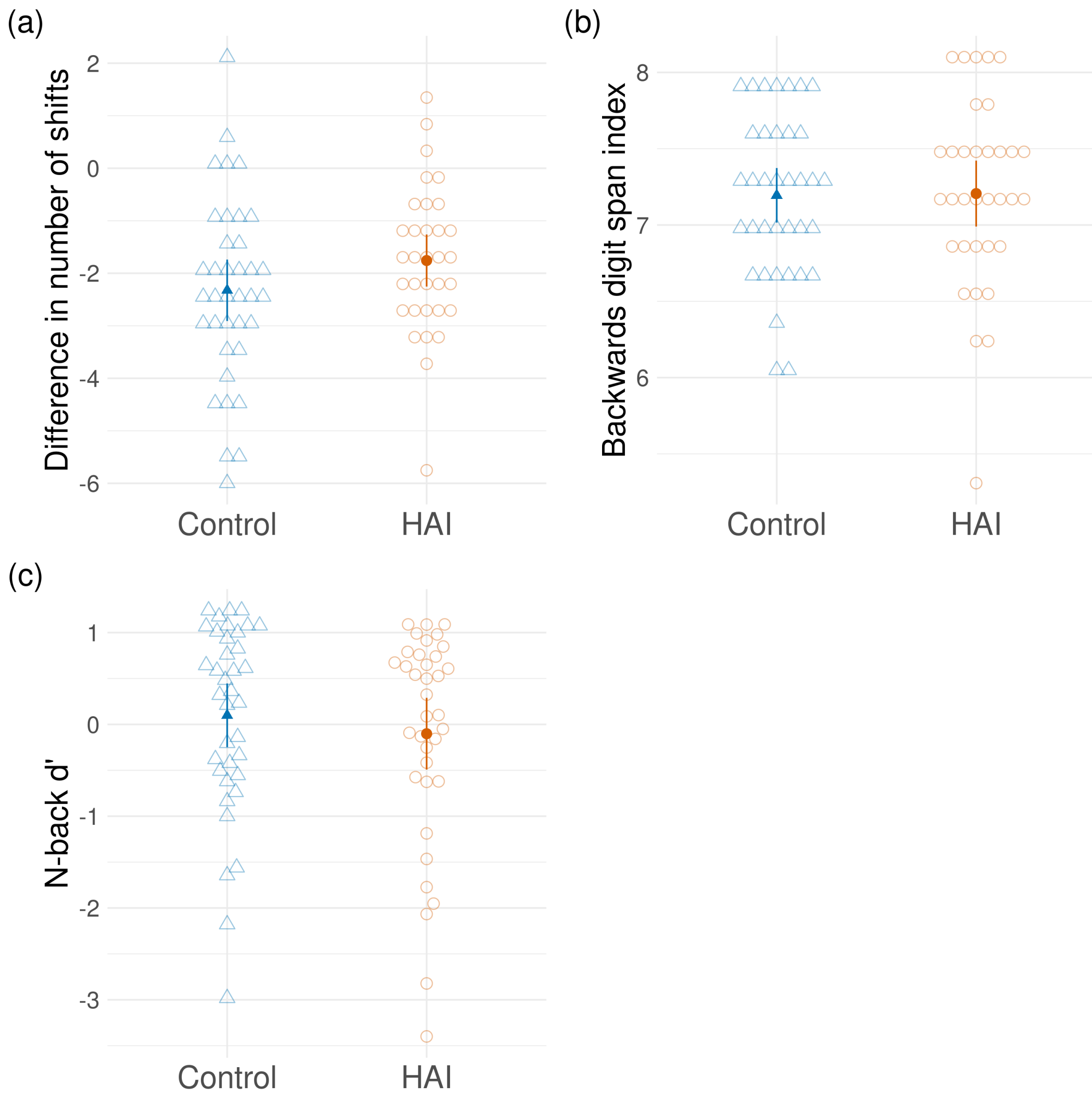

Figure S3. Post-condition predicted cognitive scores (controlling for pre-condition scores) for HAI (human-animal interaction) and control groups in Experiment 1. Scores show (a) the difference in number of attentional shifts between the two Necker cube trials, (b) the index for the backwards digit span task, and (c) d' for the n-back task. Open triangles represent individual control participant scores, open circles represent individual HAI participant scores, filled triangles and circles represent condition group means, error bars represent 95% confidence intervals.

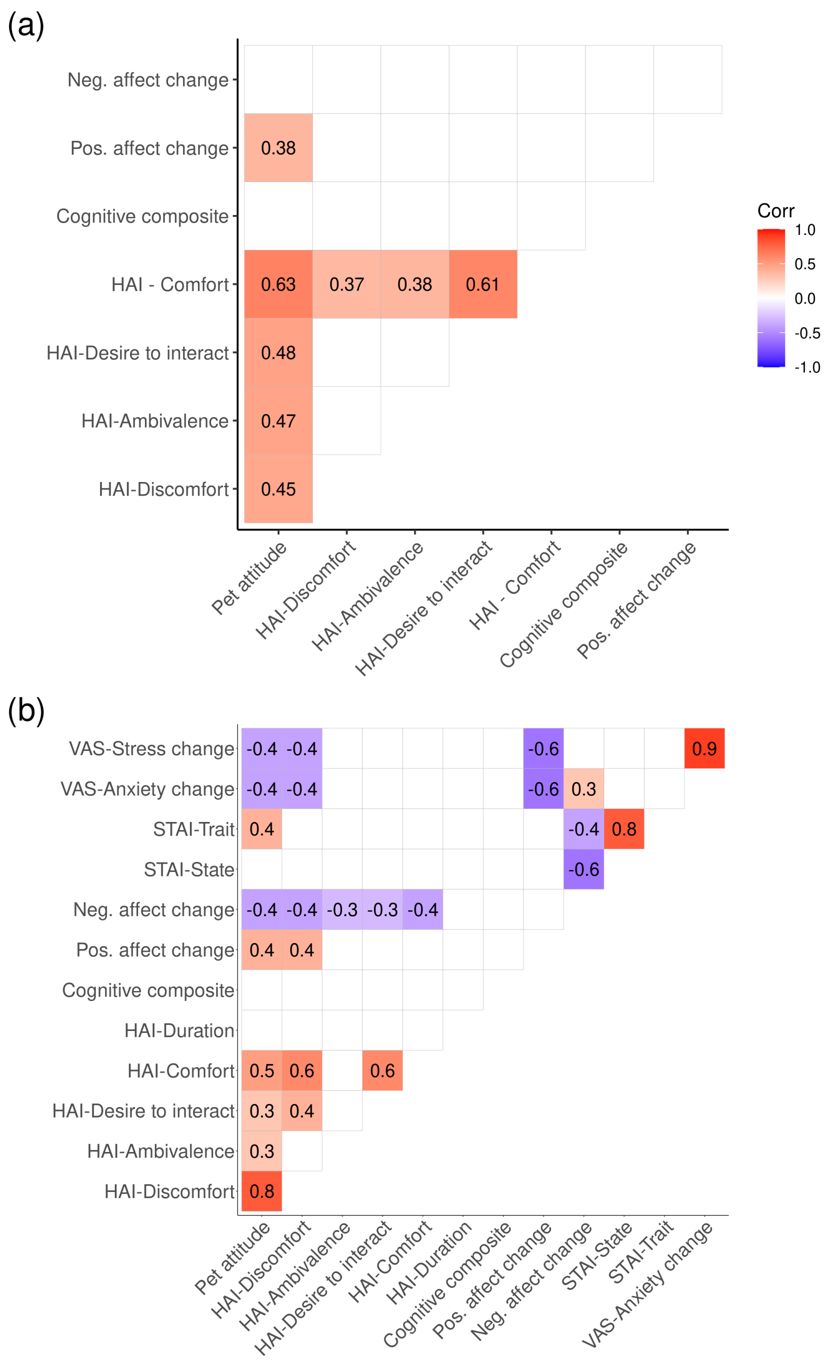

Figure S4. Animal experience correlation matrices for Experiments 1 (a) and 2 (b). Values in cells are correlation coefficients for correlations with $p < 0.05$.

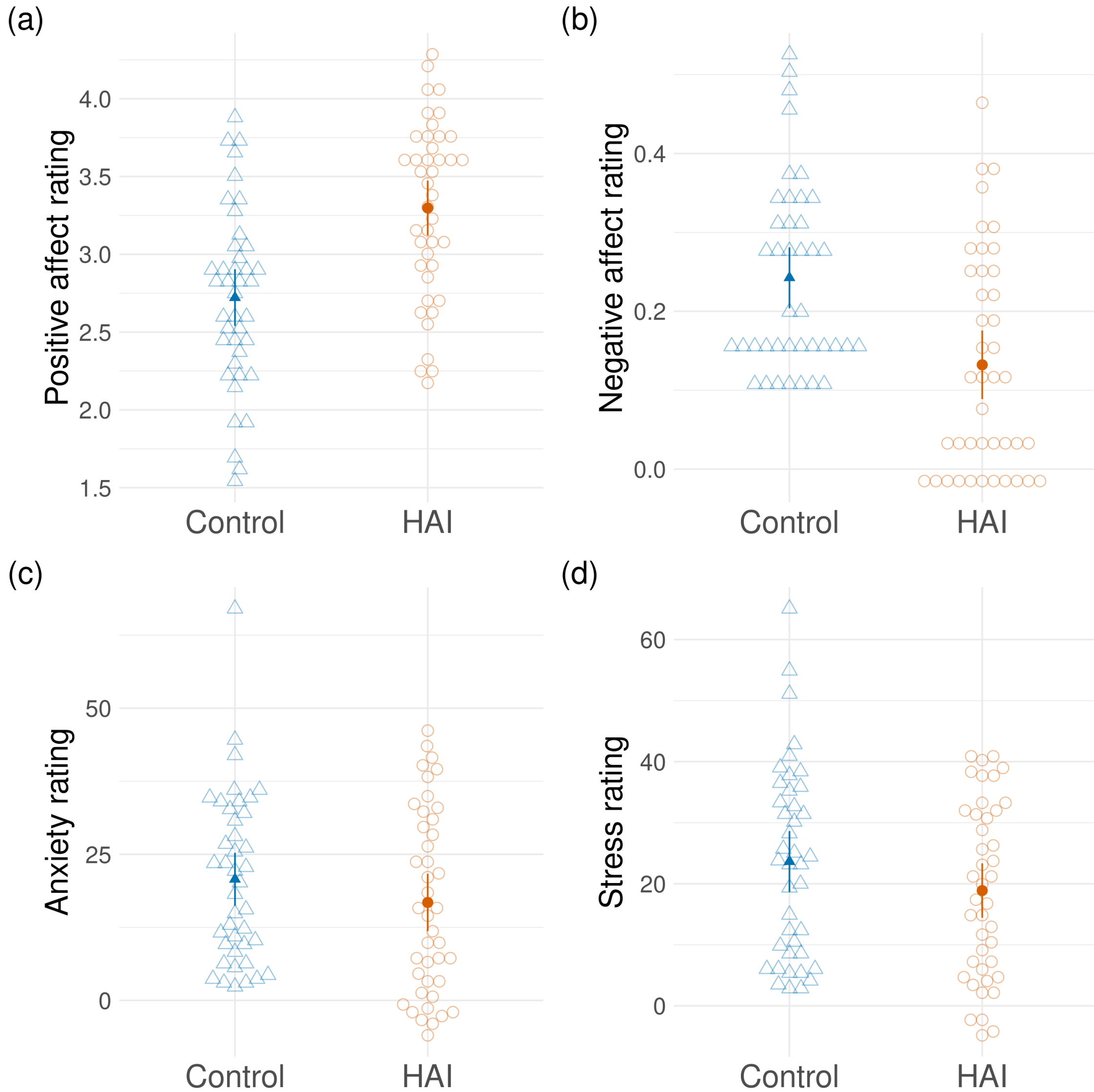

Figure S5. Post-condition predicted affect scores (controlling for pre-condition scores) for control and HAI (human-animal interaction) groups in Experiment 2. Scores show (a) positive PANAS ratings, (b) negative PANAS ratings, (c) anxiety ratings, and (d) stress ratings. Negative affect scores are log-transformed. Open triangles represent individual control participant scores, open circles represent individual HAI participant scores, filled triangles and circles represent condition group means, error bars represent 95% confidence intervals.

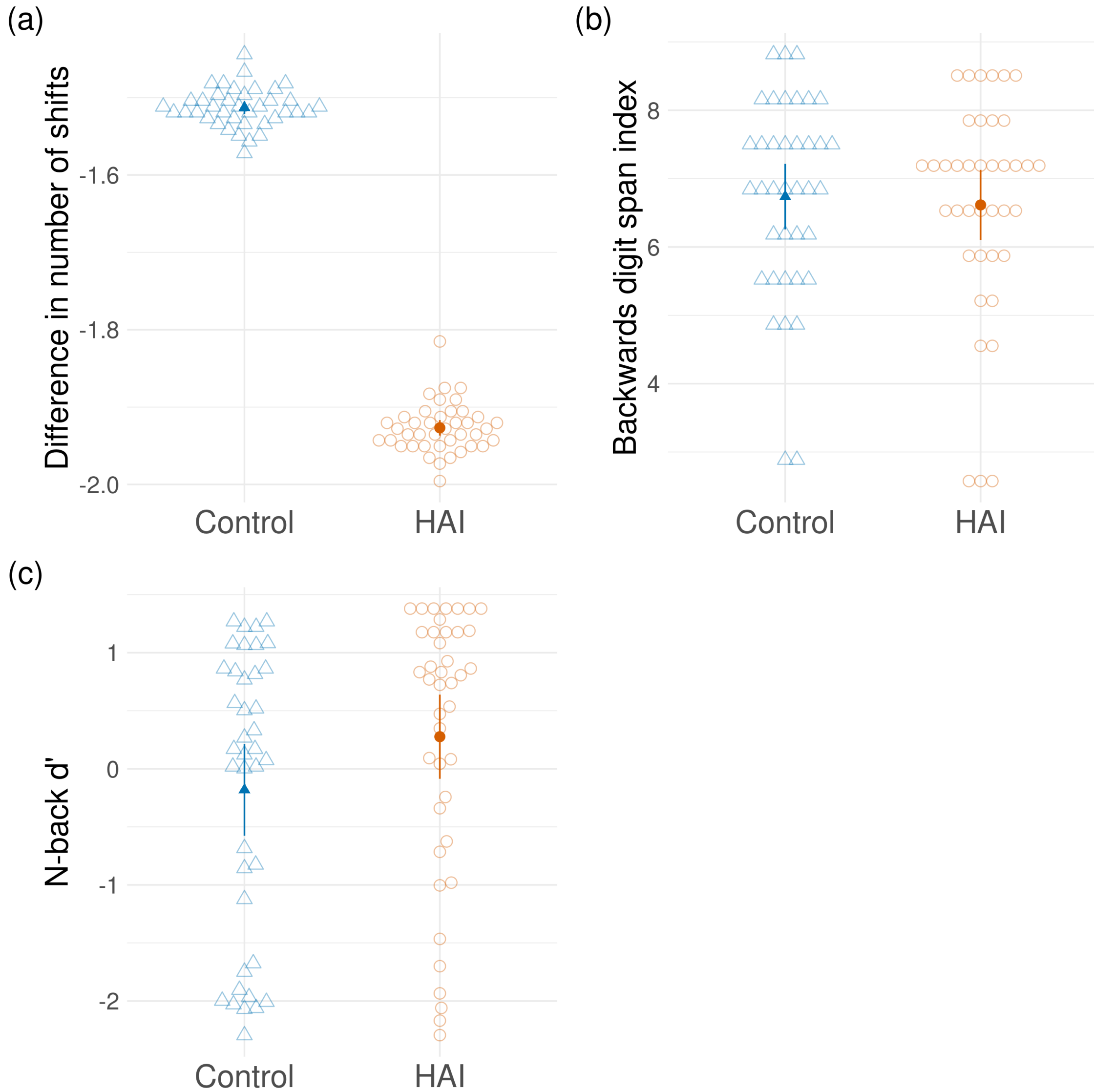

Figure S6. Post-condition predicted cognitive scores (controlling for pre-condition scores) for HAI (human-animal interaction) and control groups in Experiment 2. Scores show (a) the difference in number of attentional shifts between the two Necker cube trials, (b) the index for the backwards digit span task, and (c) d' for the n-back task. Open triangles represent individual control participant scores, open circles represent individual HAI participant scores, filled triangles and circles represent condition group means, error bars represent 95% confidence intervals.

References

Aust, F., & Barth, M. (2020). *papaja: Create APA manuscripts with R Markdown.* https://github.com/crsh/papaja

Clarke, E., & Sherrill-Mix, S. (2017). *ggbeeswarm: Categorical scatter (violin point) plots.* https://CRAN.R-project.org/package=ggbeeswarm

Empatica. (2018). *E4 Wristband User's Manual.*

Haatveit, B. C., Sundet, K., Hugdahl, K., Ueland, T., Melle, I., & Andreassen, O. A. (2010). The validity of d prime as a working memory index: Results from the "Bergen n-back task". *Journal of Clinical and Experimental Neuropsychology*, *32*(8), 871–880. https://doi.org/10.1080/13803391003596421

Harrell Jr, F. E., Charles Dupont, & others. (2020). *Hmisc: Harrell miscellaneous.* https://CRAN.R-project.org/package=Hmisc

Johnson, T. R. (2016). Violation of the homogeneity of regression slopes assumption in ANCOVA for two-group pre-post designs: Tutorial on a modified Johnson-Neyman procedure. *The Quantitative Methods for Psychology*, *12*(3), 253–263. https://doi.org/10.20982/tqmp.12.3.p253

Kassambara, A. (2019). *ggcorrplot: Visualization of a correlation matrix using 'ggplot2'.* https://CRAN.R-project.org/package=ggcorrplot

Lenth, R. V. (2016). Least-squares means: The R package lsmeans. *Journal of Statistical Software*, *69*(1), 1–33. https://doi.org/10.18637/jss.v069.i01

Mangiafico, S. (2020). *rcompanion: Functions to support extension education program evaluation.* https://CRAN.R-project.org/package=rcompanion

McCarthy, C., Pradhan, N., Redpath, C., & Adler, A. (2016). Validation of the Empatica E4 wristband. *2016 IEEE EMBS International Student Conference (ISC)*, 1–4. https://doi.org/10.1109/EMBSISC.2016.7508621

Morey, R. D., & Rouder, J. N. (2018). *BayesFactor: Computation of Bayes factors for common designs.* https://CRAN.R-project.org/package=BayesFactor

Müller, K. (2017). *here: A simpler way to find your files.* https://CRAN.R-project.org/package=here

Pedersen, T. L. (2020). *patchwork: The composer of plots.* https://CRAN.R-project.org/package=patchwork

R Core Team. (2020). *R: A language and environment for statistical computing.* R Foundation for Statistical Computing. https://www.R-project.org/

Revelle, W. (2019). *psych: Procedures for psychological, psychometric, and personality research.* Northwestern University. https://CRAN.R-project.org/package=psych

Shaffer, F., & Ginsberg, J. P. (2017). An overview of heart rate variability metrics and norms. *Frontiers in Public Health*, *5*, 258. https://doi.org/10.3389/fpubh.2017.00258 van Doorn, J., Ly, A., Marsman, M., & Wagenmakers, E.-J. (2020). Bayesian rank-based hypothesis testing for the rank sum test, the signed rank test, and Spearman's ρ. *Journal of Applied Statistics*. https://doi.org/10.1080/02664763.2019.1709053

Wagenmakers, E.-J., Love, J., Marsman, M., Jamil, T., Ly, A., Verhagen, J., Selker, R., Gronau, Q. F., Dropmann, D., Boutin, B., Meerhoff, F., Knight, P., Raj, A., Kesteren, E.-J. van, Doorn, J. van, Šmíra, M., Epskamp, S., Etz, A., Matzke, D., … Morey, R. D. (2018). Bayesian inference for psychology. Part II: Example applications with JASP. *Psychonomic Bulletin & Review*, *25*(1), 58–76. https://doi.org/10.3758/s13423-017-1323-7

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T. L., Miller, E., Bache, S. M., Müller, K., Ooms, J., Robinson, D., Seidel, D. P., Spinu, V., … Yutani, H. (2019). Welcome to the tidyverse. *Journal of Open Source Software*, *4*(43), 1686. https://doi.org/10.21105/joss.01686

Xie, Y., Allaire, J. J., & Grolemund, G. (2018). *R markdown: The definitive guide*. Chapman; Hall/CRC. https://bookdown.org/yihui/rmarkdown