Submitted: 01/21/2024
Accepted: 03/18/2024
Published: 04/05/2024

Research Article

[1]Department of Psychology, McGill University, Montréal, QC, Canada

Keywords

Replication, humility, awe, depression

Email Correspondence

alexandra.bertrand@mail.mcgill.ca

https://doi.org/10.26443/msurj.v19i1.231

Alexandra Bertrand[1], Jonah Kimmel [1], Salomé Duhamel[1], Héloïse Puel[1], Alexandra Schifano[1], and Émilie Wood[1]

A Replication Study to Evaluate the Effects of Awe on Humility

Abstract

This paper replicates Stellar and colleagues' 2018 study involving an experimental manipulation of awe using standardized video induction, as well as proposing and testing out an additional hypothesis. The previous study hypothesized that watching an awe-inducing video would lead participants to disclose fewer strengths when writing about their self-perceived strengths and weaknesses. In addition to the replication, we hypothesized that participants with higher scores of depressive symptomatology (DS) would list fewer strengths due to diminished self-concept and self-efficacy. Ninety-four undergraduate psychology students were recruited from McGill University ranging from 18 to 35 years of age. Participants were randomly assigned to either the awe-inducing or neutral video condition and then filled out measures of humility, emotional reactions, and DS. In contradiction with the original study, participants in the awe condition and the neutral condition did not significantly differ in their ratio of disclosed strengths to weaknesses, therefore no significant correlations were found between awe and humility or humility and depression. Additionally, results indicated that participants with greater DS did not list fewer strengths than those scoring lower on the Center for Epidemiologic Studies Depression Scale (CES-D). We were unable to directly replicate the original study and thus rejected our alternate hypothesis. This study had various potential limitations, among which are the possibilities of self-report bias, issues regarding convenience sampling, and bias due to time constraints. The current study advances the literature by examining depression as it relates to awe and humility. Further research is needed to differentiate lab-induced awe from natural experiences of awe and to identify possible moderating factors on humility.

Introduction

According to Lee & Ashton (2005, in Stellar et al., 2018), humility involves a stable, sincere, and non-restrictive view of one's self, in conjunction with a recognition of the value of things external to the self[1]. Humility is a foundational virtue that acts as a counterforce to self-centered tendencies like entitlement, arrogance, and narcissism, and is essential to living in social groups[1]. Despite this significance, little is known about how this virtue can be bolstered[1]. To address this gap, Stellar and colleagues posited that humility could be influenced by experiences of awe, such as encounters with forces or situations that are immeasurable and cerebrally profound. And through this encounter, awe has the capacity to induce a dramatic shift in an individual's self-concept, characterized by a reduced sense of self[1]. In other words, experiencing awe would generate greater humility. Across five studies, Stellar and colleagues aimed to explore the connection between awe and humility, assessing whether momentary experience of awe could promote humility by using an appraisal-tendency framework to clarify the underlying process behind this effect[1].

This paper focuses on a replication of Stellar and colleagues' third study, which involved an experimental manipulation of awe using standardized video induction, as well as proposing and testing out a new additional hypothesis[1]. In the original study, humility was evaluated as a behavioural measure, such that greater humility was operationalized as a more balanced presentation of one's own strengths and weaknesses to other people, such as less bragging or self-promotion[1]. Furthermore, the hypothesis for this investigation posited that, in contrast to a neutral control condition (i.e., a video about fence building), momentary experiences of awe induced by a video illustrating the universe expanding would lead to a more equitable disclosure of personal strengths and weaknesses[1]. This hypothesis is important as it could be a first step towards eventually using inductions of awe to promote humility in people.

In alignment with their aforementioned hypothesis, the induction of awe prompted participants to exhibit greater humility, as manifested in their disclosure of fewer personal strengths than participants in the neutral condition[1]. However, there was no effect of condition on the number of weaknesses listed by participants[1]. Given that the experimental condition evoked more happiness and awe, a multiple regression analysis was conducted using awe and happiness as predictors of humility[1]. The original results revealed that neither awe nor happiness were significant predictors, indicating the inability to establish the distinctive impact of awe, as self-reports of this emotion did not forecast levels of humility[1].

Additional Hypothesis

A recent systematic review found that the mean prevalence of depression in university students to be 30.6%, higher than the mean of 9% found in the general population (SD=3%). Thus, due to the inconclusive findings of the original study, as well as the results from the systematic review regarding our participant sample, we hypothesize that depression levels might mediate humility (more so than awe). Accordingly, we consider the possibility that depression, as a potential confound, could have influenced participants' written self-evaluation, since having a negative view of oneself, or a negative self-schema, has been reliably linked with depressive symptomatology (DS)[2–5]. Hence, our proposed alternative hypothesis is that participants with greater depressive symptomatology, as measured by higher scores on the Center for Epidemiologic Studies Depression Scale (CES-D), will report fewer strengths and thus greater humility than participants with less DS, in both control and experimental conditions. To test this alternate hypothesis, we first replicated the original third study by Stellar and col-

leagues and then had participants fill out the CES-D. Finally, our hypothesis is relevant to the original conclusions of Stellar et al. since we believe listing fewer strengths may be mediated by self-diminishment that is linked to participants' DS rather than humility, since self-diminishment in depression is thought to be related to negative self-concept, whereas humility is usually concerned with a more neutral view of yourself, a lack of pride, or not considering yourself to be "better than others".

Methods

Participants

We attempted to recruit 104 participants for the study, as the original study we were replicating, Study 3 of Stellar et al. (2018), had 104 participants recruited online, with a final sample of 85 participants. Recruitment was conducted through Sona Systems, a research platform affiliated with McGill University in which participants were presented with a general outline of the study and its various requirements. Due to limitations in the student study pool, our participant sample consisted of 94 undergraduate psychology students recruited from McGill University ranging from 18 to 35 years of age. Participant gender and age were not self-reported during the study procedure, so we are unable to make any inferences about differences between ours and the original study's demographic samples. The participants voluntarily chose to participate for 0.5% psychology course credit as an incentive, which could be applied towards a psychology course that they were registered in. We were one of two research teams concurrently replicating the Stellar et al. (2018) study, therefore participation restrictions were implemented as participants who took part in this study could not participate in the other replication of the original study, and vice versa, to reduce validity issues.

Procedure

Data was collected from November 2nd to November 17th, 2023. Participants arrived at the lab and were provided with brief instructions on where to sit and how to proceed with the study. Using a between-subjects posttest experimental design, participants were randomly assigned to either the experimental (awe) or control (neutral) conditions. Participants in the awe condition watched a two-minute video meant to depict the expansiveness of the universe as it slowly zoomed out from planet Earth and exposed more and more of the universe. Participants in the neutral condition watched a two-minute instructional video on how to build a fence. In previous mood-induction-based studies, this video was not shown to elicit strong emotions in participants other than relaxation and calm[6].

Measures

Participants were instructed to watch a two-minute video viewed on a desktop computer screen in a university computer laboratory. As multiple participants completed the study in the lab at the same time, alternating computers were used to seat participants, allowing for a minimum distance of 2 meters between them. Participants were also provided with headphones to hear the audio from the video. The same videos from YouTube used by the original study were used in this replication study. Once the video ended, they were automatically directed to a new page of the survey where they were instructed to write about their strengths, followed by their weaknesses, for two minutes. They were instructed to write as though they were discussing their qualities with a person they just met, to standardize the response type across participants as well as augment the interactive nature of the online setting to be more in line with the other 4 studies of the original paper, all of which were conducted in-person[1]. A timer was visible to participants to show them how much time they had left. Once the timer ended, the survey automatically advanced to a new page where the participants were instructed to report how much awe, happiness, fear, wonder, and amazement they felt while watching the video. They scored their feelings on a scale of 1 (not at all) to 7 (very much). They were then asked questions meant to address their comprehension and completion of this portion of the study, such as explaining what the video was about, what they believed the study to be about, and if the video loaded properly and played all the way through.

Participants were then prompted to complete the Center for Epidemiologic Studies Depression Scale (CES-D) to measure their level of depressive symptomatology over the last week. The CES-D is a self-report scale measuring depressive symptoms that is meant to be used in the general population. The measure consists of questions selected to identify how often participants experienced depressive symptoms in the past week[7]. Participants were presented with statements such as "I felt fearful", "I was happy", "I was bothered by things that usually don't bother me", and "I enjoyed life" and were asked to indicate how often they felt such a way in the last week using a Likert Scale from *Rarely or none of the time (less than 1 day) to Most or all of the time (3-4 days)*. Higher scores on the measure indicated more depressive symptoms, with scores of 16 or higher being indicative of a risk for clinical depression[7]. Though various shortened versions exist (for example, a 10-item version), the complete 20-item scale was used on participants (See Appendix A) as it only requires an average of five minutes to complete by the general population who do not suffer from cognitive impairments[8].

Consistent with Stellar et al., once data was collected, two coders were assigned to read participants' responses and count the number of strengths and weaknesses stated[1]. The original study did not make any mention of what coding rules were followed by raters during the analysis. Thus, prior to the start of coding, a codebook was established. Various ideas and issues were established prior to coding and subsequently translated into formal rules that coders were to follow while rating the number of strengths and weaknesses written by participants (See Appendix B). Coders were strictly informed not to communicate once coding began. Discussion about participants' listed strengths and weaknesses was to be avoided both between and within groups of raters to ensure no rater influenced another and biased the results. The average of the two coders was taken and considered as the number of strengths and weaknesses listed by participants. Since participant strengths and weaknesses were written free-form and coded by rater pairs, interrater reliability was examined for each rater pair on raw strength and weakness counts. Intraclass correlation coefficients measuring agreement between rater pairs showed excellent reliability or acceptable agreement for all examined response pairs save one, with the divergent pair showing moderate reliability and tentatively acceptable agreement, per generally accepted threshold values. Further examination of the divergent response pair showed one problematic observation. As there was no theoretical or pre-registered reason for modifying these observations per the code book or previous literature, all data were retained for analysis. See Appendix C for reliability ratings.

Results

Ninety-four participants completed the study. It is important to note that a significant number of participants listed zero weaknesses, which led to them being excluded from certain computations due to missingness resulting from subsequent log-transformations. This participant attrition resulting from analysis may impact the validity of our original study and is discussed further in the next section. Similar to the replicated study, participants described several strengths and weaknesses that were positively skewed, so these variables were log transformed. Per the procedure of the

previous study, the number of strengths was divided by the number of weaknesses to arrive at a humility score for each participant[1]. The resulting scores were also not normally distributed; therefore, they too were log transformed. All available participant data was used to run the analyses.

The analyses showed that participants in the awe and the neutral condition did not significantly differ in their balance between strengths and weaknesses disclosed (i.e., their humility score, $t(69) = -0.34$, $p = 0.438$); see Appendix D for more information. This contradicts the original study, where a significant difference was found, $t(84) = 2.38$, $p = 0.02$.

Next, strengths and weaknesses were examined individually, which is where the original study found that participants in the awe condition reported significantly fewer strengths and a similar number of weaknesses. In the current study, participants in the awe condition reported slightly more strengths ($t(91) = 0.85$, $p = 0.397$) and slightly more weaknesses ($t(82) = 0.50$, $p = 0.618$), and both those differences were not significant when compared to the neutral condition; see Appendix E for more information. To determine the overall effect of the awe manipulation on the variable of humility, the log-transformed variable of humility was regressed on reported feelings of awe. There was no significant difference between awe and neutral groups, $F(1, 69) = 0.11$, $p = 0.738$.

To ensure that the manipulation was successful in inducing awe in the participants in the experimental condition, participant-reported feelings of awe were regressed onto their condition. Results indicate that the experimental condition was a significant positive predictor of awe scores, $b_1 = 2.85$, $t(91) = 8.80$, $p < 0.001$. The average reported awe for participants in the neutral condition was 2.13, $t(91) = 9.33$, $p < 0.001$. Participant condition accounted for $R^2_{\text{adjusted}} = 45.4\%$ of the variance in reported awe. Similar results were found when the same calculations were performed with a composite of awe that included reported feelings of awe, amazement, fear, and wonder. Experimental condition was a significant predictor of this awe composite score, $b_1 = 2.55$, $t(91) = 12.10$, $p < 0.001$, and participant condition accounted for $R^2_{\text{adjusted}} = 61.25\%$ of the variance. These results confirm that awe was significantly different between conditions and disparage the manipulation failure as a potential explanation for a failure to replicate the results of the original study. See Appendix F for a figure of these results.

In addition to the replication, it was alternatively hypothesized that participants with higher scores of DS would list fewer strengths due to a diminished but non-negative self-concept and self-efficacy. Appendix G shows a figure with respondents plotted by their total score on the CES-D and the number of strengths reported. Participants wrote 4.8 strengths on average, and depression scores were 22.6, with a standard deviation of 10.7, which is noteworthy in that the average participant would score as being at moderate risk of DS per the approved cutoffs. Greater DS, as indicated by higher CES-D scores, did not significantly predict the number of strengths reported, $b_1 = -0.01$, $t(91) = -0.73$, $p = 0.470$, and CES-D scores accounted for $R^2_{\text{adjusted}} < 0.1$ % of the variance.

Discussion

Stellar et al. conducted a study to investigate if there was an association between the feeling of awe and humility[1]. The study investigated whether the feeling of awe would lead to humility through watching a video and then listing personal strengths and weaknesses. They hypothesized that watching an awe-inducing video would lead participants to write fewer strengths before writing their weaknesses. Their results demonstrated that their hypothesis was correct, and they found a significant association between awe and humility. This led the researchers to conclude that awe does in fact induce humility[1].

The goal of the current study was to determine whether these findings could be replicated or if there was potentially another factor that could have affected their results. We conducted a direct replication of the original study and included a potential confound of depression[1]. We speculated that the original study's findings, wherein the participants in the experimental group listed fewer strengths but the same number of weaknesses, could have been mediated by depression levels. Beck's cognitive model of depression has found that depressive symptoms are a result of the negative cognitive triad, meaning that people with depression are more prone to negative self-schemas and negative self-thought[2]. Due to the original study's operationalization of humility as listing strengths and weaknesses, we hypothesized that listing fewer strengths could have been related to negative self-schemas, as depressed individuals often find it hard to see the positive in life, in our case, perhaps their own strengths. However, based on our findings, no significant correlations were found between awe and humility, or depression and humility, leading to both the rejection of our alternate hypothesis and the failure to directly replicate the original study. Specifically, our results show that participants with greater DS did not list fewer strengths compared to those who scored lower on the CES-D. Therefore, our study's findings seem inconsistent with existing literature on depression and Beck's cognitive triad, as having a negative view of oneself, or negative self-schema, has been reliably linked with DS[2–5]. Another study that evaluated depression in adolescents found that those with a more positive self-concept appeared to have less DS[9]. Therefore, according to this myriad of studies, the written self-evaluations used to measure humility should have been affected negatively in those that had higher depression. According to a systematic review, depression is a highly prevalent problem in university students that causes a decrease in self-esteem[10], demonstrating once again that our results should have shown a decrease in strengths in participants with a higher score for depression.

Moreover, our study has not found any mediating effects in the original study's association between awe and humility. In our study, the additional confound of depression was not found to explain the association, directly or as a mediating variable, between awe and humility per the results from the replicated study.

Limitations

The study we conducted was limited with regards to the internal validity of its conclusions, the external validity, and the power of the study's entirety. First, awe is an extremely vast and abstract concept, often considered to be a rare and intense emotion[11], therefore our manipulation of evoking it through a short YouTube video should be more closely examined. The post-test-only measure of awe induction in participants allows for the possibility that the difference observed between the experimental and control groups could be explained by lower levels of awe in the control group, rather than the assumed heightened levels in the experimental group. The neutral fence-building video used in the control group was assessed in the original study for emotions elicited by the video as outlined in our methods, but not whether it diminished other emotions[1]. This allows for the possibility that the control video had an effect, such as lowering levels of awe, wonder, or amazement, rather than being truly neutral. This potential confound could mean that the experimental video did not induce awe in the first place, impacting the validity of our conclusions. To fix this manipulation issue in the future, we could either pre- and post-test the neutral video to observe potential awe-depleting effects, or simply use a more valid and reliable awe-induction technique as our manipulation, discussed later in the future directions section. Similarly, the original study did not test the validity of their operationalization of humility as a ratio of self-reported strengths to weaknesses[1]. Thus, we cannot be confident in the construct validity of the results gathered through these measures, both subjectively (face and content validity) and objectively (criterion, convergent, and discriminant validity).

It is possible there were order effects in both humility and awe induction measurements. This would be due to all participants being told to list strengths before moving on to weaknesses, as well as the induced emotion questionnaire always beginning with self-reported awe. The two-minute time limit appeared to have caused issues with a proportion of the participants, as some were cut off either before or while they listed weaknesses. These issues skewed the resulting ratios and potentially impacted the measure's ability to capture humility. If one were to conduct this study again, they would benefit from using a combination of measurements; for example, adding reliable physiological measurements of awe such as skin conductance and pupil diameter[12], among other self-report measures of humility[13]. The combination of self-report and physiological assessment has been found to have higher validity than either one alone[11]. If the self-report measure of humility was retained, we could reduce order effects by removing the timer and by counterbalancing the order of strengths & weaknesses across participants, as well as randomizing the order of the emotions listed in the awe-induction questionnaire. Similarly, we could use a different measure to record awe induction, such as the Situational Awe Scale, a valid and reliable measure of awe induction in the lab[14], rather than having participants self-report the level of awe and other emotions they felt, as this method is vulnerable to diverse biases.

A limitation concerning the external validity of our study was selection bias, given our convenience sample of younger-skewing undergraduate psychology students. Our sample consisting solely of psychology students presents a serious limitation, as this population is taught about the use of deception in psychological experiments and thus may be more suspicious and critical of studies in which they participate, as demonstrated through the insightful correct guesses of our hypothesis from some participants. This could have led to demand characteristic biases, wherein participants distort their responses, whether consciously or unconsciously, in accordance with the presumed hypothesis. Thus, viewing the experimental awe video and then immediately being asked to report awe levels may have had a suggestive effect on the participants' responses and led to biased responses. Another possible validity issue is the presence of self-report bias due to social desirability effects, reflected in higher-than-average levels of trait agreeableness in psychology undergraduates who volunteer for studies[15]. Self-report answers can be strongly influenced (consciously or not) by how participants want others to perceive them, leading to biased responses; in this case potentially leading participants to list more strengths than they would have had it been a completely anonymous online study, like the original. The computer-based and single-session nature of the current study could have protected against this marginally, but it nonetheless may have had an effect. Reactivity bias, due to the presence of researchers, poses another potential validity issue. All these biases can limit the generalizability of conclusions, as the results are not necessarily representative of the broader population in real-life situations. Going forward, we would attempt to increase our external validity by using a more representative sample, which could potentially reduce threats such as demand characteristics and volunteer bias, and hopefully increase the generalizability of results.

The suggestions presented for reducing order effect bias, such as randomization and counterbalancing, could potentially reduce demand characteristics as well, as the study's focus would be less obvious to participants. To reduce reactivity bias, we could move the study online, although that would impact the potential of using physiological measurements. However, to reduce the risk of confounds, we intentionally placed participants at computers with a space between each participant, when space permitted.

Another limitation of our study is the small sample size: coupled with the small and statistically insignificant effects we found, it is likely that the study had low power, increasing the likelihood of making an error regarding the conclusions of the study—in this case failing to reject the null hypothesis.

Lastly, since certain participants failed to report weaknesses, but not strengths, the number of excluded observations varied across calculations. Computations including ratios or multiple log transformations specifically required that more observations be excluded from the analyses, because mathematically, dividing by zero is undefined. This represented an important mathematical limitation, leading our statistical analyses to be prejudiced against participants who did not write weaknesses for several reasons such as time limit, reactivity, etc. It is conceivable that taking those participants out of the analyses could have skewed the data in a problematic way. In future research, listing no weaknesses would be added to the exclusion criteria and considered as a failure to follow instructions to help minimize variability in sample size across calculations. We would also recruit a sufficient sample size with responses of both strengths and weaknesses to obtain enough power to detect small effects. Furthermore, developing, validating, and assessing participants with a measure of assessing humility by self-report of strengths and weaknesses that does not exclude participants who report no strengths or weaknesses will be important in strengthening the validity of studies that measure this construct, through reduced attrition of participants.

Future Directions

The study of awe and humility comes with many unanswered questions that future directions of exploration may help to answer. One study, for example, discovered a potential mediating effect of dispositional humility on DS, as a self-regulating mechanism involved in self-differentiation[16]. Thus, there does seem to be a correlation between depression and humility, although perhaps more complex than we believed and not directly related to negative self-image, possibly explaining why we did not find a significant correlation between DS and fewer strengths being listed in our additional hypothesis. Another study found that humility in university students was positively associated with conscientiousness, openness, love of life, happiness, and self-efficacy, but negatively associated with agreeableness, neuroticism, and depression[17]. However, as stated previously, university students who volunteer for psychology studies tend to have relatively higher levels of agreeableness[15], which correlates negatively with humility. Another study found a negative correlation between humility, specifically the H dimension of the HEXACO personality model, and a wide array of Machiavellian behaviours such as criminal activity and power-seeking tendencies, potentially hinting at a protective social benefit related to humility[18] in the form of stronger hierarchical structures[19]. Namely, leaders with higher state and/or trait humility are less likely to make decisions that are self-serving at the expense of the group[18], and this reinforces follower humility while acting as a protective factor against Machiavellianism in followers. This would, in turn, reduce their potentially detrimental power-seeking tendencies.[19] Cross-cultural studies are of interest, as recent investigations into the reproducibility of the Humility-honesty dimension as a personality factor has not shown successful retrieval by Exploratory Structural Equation Modeling across all cross-cultural groups[20]. Research by Ion and colleagues (2017) was unable to reproduce a model that included additional predictive validity from a 6th facet of personality from the HEXACO model[20], Humility-honesty, compared to the five-factor model by Costa and McCrae[21], for certain cultural groups. Examining differences in the experience and expression of humility based on cultural differences, such as socioeconomic status, previous experience with awe, or predominance of differing majority religions, may help us understand the etiology of humility within specific cultures. Future research could attempt to clarify the tangled relationships between experiences of awe, trait and state humility, self-schema, DS, as well as potential moderation through demographic factors and personality traits.

One final avenue of future research involves studying how lab-induced awe differs from natural experiences of awe in the real world, and whether humility is more reliably induced by authentic experiences of awe. Chirico et al. (2017) found that since awe is such a complex emotion, with experi-

ences described as meaningful and profound, the level of awe induced in the lab is heavily dependent on the immersive quality of the induction[11]. The results of their study concluded that immersive videos, considered to be the most realistic form of virtual reality (VR), significantly enhanced self-reported (subjective) and physiological (objective) measurements of awe when compared to 2D screen videos, like the one we used in our study[11]. VR accessibility and quality are rapidly increasing with each passing year of technological advancements, proposing an immensely exciting solution to the future of experimental research, in our case specifically relating to stronger and more reliable manipulation of awe in laboratory settings.

Conclusion

While our results fail to reproduce the findings of the previous paper[1], the future directions and conceptual replications of this group of studies will further examine the link between awe and humility and how they contribute to a uniquely human condition. To this end, this field of research remains invaluable in examining experiences that until recently have been the purvey of only philosophy and religion.

Acknowledgements

We want to thank Dr. Eric Hehman, Assistant Professor at McGill University, for his guidance and support throughout the project's development, execution, and final paper composition.

References

1. Stellar, J. E. & et al. Awe and humility. *J. Pers. Soc. Psychol.* **114,** 258–269 (2018). https://doi.org/10.1037/pspi0000109
2. Beck, A. T. & et al. Levels of hopelessness in DSM-III disorders: A partial test of content specificity in depression. *Cognitive Ther. Res.* **12,** 459–469 (1988). https://doi.org/10.1007/BF01173413
3. Mahali, S. C., Beshai, S., Feeney, J. R. & Mishra, S. Associations of negative cognitions, emotional regulation, and depression symptoms across four continents: international support for the cognitive model of depression. *B.M.C. Psychiatry* **10,** 18 (2020). https://doi.org/10.1186/s12888-019-2423-x
4. Phillips, W. J., Hine, D. W. & Thorsteinsson, E. B. Implicit cognition and depression: a meta-analysis. *Clin. Psychol. Rev.* **30,** 691–709 (2010). https://doi.org/10.1016/j.cpr.2010.05.002
5. Sowislo, J. F. & Orth, U. Does low self-esteem predict depression and anxiety? A meta-analysis of longitudinal studies. *Psychol. Bull.* **139,** 213–240 (2013). https://doi.org/10.1037/a0028931
6. Stellar, J. E., Manzo, V. M., Kraus, M. W. & Keltner, D. Class and compassion: socioeconomic factors predict responses to suffering. *Emotion* **12,** 449–459 (2012). http://dx.doi.org/10.1037/a0026508
7. Radloff, L. S. The CES-D scale: A self-report depression scale for research in the general population. *Appl. Psychol. Meas.* **1,** 385–401 (1977). https://doi.org/10.1177/014662167700100306
8. Shaffer, K. in *Encyclopedia of Quality of Life and Well-Being Research* (ed Michalos, A. C.) 637–640 (Springer Netherlands, 2014).
9. Hards, E. et al. Self-evaluation and depression in adolescents with a chronic illness: a systematic review. *Clin. Child Psychol. Psychiatry* **28,** 382–397 (2023). https://doi-org/10.1177/13591045221115287
10. Ibrahim, A. K., Kelly, S. J., Adams, C. E. & Glazebrook, C. A. A systematic review of studies of depression prevalence in university students. *J. Psychiatr. Res.* **47,** 391–400 (2013). https://doi.org/10.1016/j.jpsychires.2012.11.015
11. Chirico, A. & et al. Effectiveness of immersive videos in inducing awe: an experimental study. *Sci. Rep.* **7,** 1218 (2017). https://doi.org/10.1038/s41598-017-01242-0
12. Takano, R. & Nomura, M. A closer look at the time course of bodily responses to awe experiences. *Sci. Rep.* **13,** 22506 (2023). https://doi.org/10.1038/s41598-023-49681-2
13. Nielson, R. & Marrone, J. Humility: our current understanding of the construct and its role in organizations. *Int. J. Manag. Rev.* **20** (2018). https://doi.org/10.1111/ijmr.12160
14. Krenzer, W. L. D., Krogh-Jespersen, S., Greenslit, J., Price, C. A. & Quinn, K. A. Assessing the experience of awe: validating the situational awe scale. Preprint at: https://doi.org/10.31234/osf.io/dsytn (2020).
15. Dollinger, S. J. & Leong, F. T. L. olunteer bias and the five-factor model. *J. Psychol.* **127,** 29–36 (1993). https://doi.org/10.1080/00223980.1993.9915540
16. Jankowski, P. J., Sandage, S. J. & Hill, P. C. Differentiation-based models of forgivingness, mental health and social justice commitment: mediator effects for differentiation of self and humility. *J. Posit. Psychol.* **8,** 412–424 (2013). https://doi.org/10.1080/17439760.2013.820337
17. Ross, L. T. & Wright, J. C. Humility, personality, and psychological functioning. *Psychol. Rep.* **126,** 688–711 (2023). https://doi.org/10.1177/00332941211062819
18. Ashton, M. C. & Lee, K. The HEXACO model of personality structure and the importance of the H factor. *Soc. Personal. Psychol.* **2,** 1952–1962 (2008). https://doi.org/10.1111/j.1751-9004.2008.00134.x
19. Shu-Chen, C., Wen-Qian, Z. & Na-Ting, L. Leader humility and machiavellianism: investigating the effects on followers' self-interested and prosocial behaviors. *Front. Psychol.* **13,** 742546 (2022). https://doi.org/10.3389/fpsyg.2022.742546
20. Ion, A. & et al. A cross-cultural analysis of personality structure through the lens of the HEXACO model. *J. Pers. Assess.* **99,** 25–34 (2017). https://doi.org/10.1080/00223891.2016.1187155
21. McCrae, R. R. & John, O. P. An introduction to the five-factor model and its applications. *J. Pers.* **60,** 175–215 (1992). https://doi-org.proxy3.library.mcgill.ca/10.1111/j.1467-6494.1992.tb00970.x
22. Krippendorff, K. *Content Analysis: An Introduction to its Methodology* (Sage publications, 2018).

Appendix A

Table 1. Center for Epidemiologic Studies Depression (CES-D) Scale. Instructions: Please read each question carefully, then circle one of the numbers to the right to indicate how you felt or behaved during the past week, including today[7]. Note: scores range from 0 to 60. Higher scores indicate more depressive symptomatology.

During the past week:	Rarely or none of the time (less than 1 day)	Some or a little of the time (1-2 days)	Occasionally or a moderate amount of the time (3-4 days)	Most or all of the time (5-7 days)
1) I was bothered by things that usually don't bother me	0	1	2	3
2) I did not feel like eating; my appetite was poor	0	1	2	3
3) I felt that I could not shake off the blues even with help from my family and friends	0	1	2	3
4) I felt that I was just as good as other people	3	2	1	0
5) I had trouble keeping my mind on what I was doing	0	1	2	3
6) I felt depressed	0	1	2	3
7) I felt that everything I did was an effort	0	1	2	3
8) I felt hopeful about the future	3	2	1	0
9) I thought my life been a failure	0	1	2	3
10) I felt fearful	0	1	2	3
11) My sleep was restless	0	1	2	3
12) I was happy	3	2	1	0
13) I talked less than usual	0	1	2	3
14) I felt lonely	0	1	2	3
15) People were unfriendly	0	1	2	3
16) I enjoyed life	3	2	1	0
17) I had crying spells	0	1	2	3
18) I felt sad	0	1	2	3
19) I felt that people disliked me	0	1	2	3
20) I could not get "going"	0	1	2	3

Appendix B

Codebook Instructions for Counting Participant-Reported Strengths and Weaknesses

A codebook was established prior to the start of coding. If participants specifically stated a trait as a strength/weakness, coders were to count them as such regardless of how the coder viewed the trait. Some participants may follow a trait by a description of such a trait. This was only to be counted as one strength/weakness. Raters were to assign participants positive integers with no decimal values. If a participant listed strengths/weaknesses that may seem similar to the rater, they were to be counted as separate traits. As a time limit was set for participants while writing their strengths and weaknesses, it is possible that they may have been cut off while writing as the survey moved on to the next section. Raters were to count these partially stated traits provided that they made sense but were not to assume the rest of the sentence.

Appendix C

Table 2. Table of Interrater Reliability Analyses by Rater Pairs. ICC refers to interrater reliability calculated through two-way mixed effect average measure intraclass correlation coefficient measuring absolute agreement. α-Reliability refers to Krippendorff's alpha, a coefficient designed to measure agreement between observers. All analyses were performed using R Statistical Software version 4.3.1 (R Core Team 2023) via the irr R package v0.84.1. ICC scores of 0.5 to 0.75 (*) and greater than 0.9 (**) are indicative of moderate and excellent reliability, respectively[22]. Alpha scores between 0.66 and 0.80 (*) and between 0.81 and 0.99 (**) indicate tentatively acceptable and acceptable agreement, respectively[7]

Rater Pair	*n* Observations	Variable	ICC	*a*-Reliability
Pair 1	29	Strengths	0.72*	0.69*
		Weaknesses	0.91**	0.94**
Pair 2	32	Strengths	0.97**	0.94**
		Weaknesses	0.98**	0.99**
Pair 3	32	Strengths	0.97**	0.95**
		Weaknesses	0.95**	0.97**

Appendix D

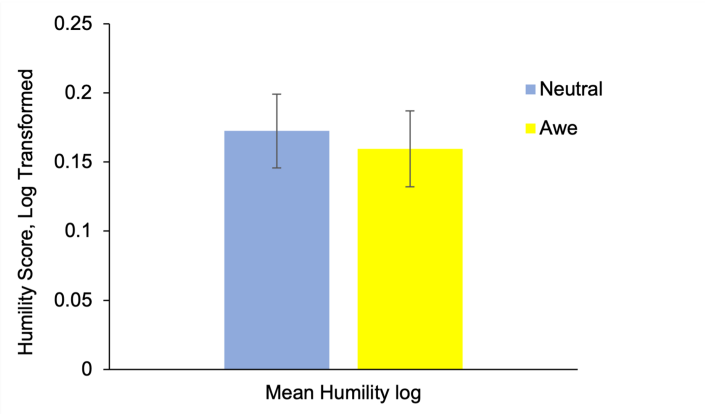

Figure 1. Calculated Humility Scores by Condition. Humility is calculated by dividing log-transformed average strength scores by log transformed weakness scores (error bars show standard errors). Humility scores in the awe condition (M = 0.16, SD = 0.16) when compared to the neutral condition (M = 0.17, SD = 0.16) were not significantly different $t(69) = -0.34$, $p = 0.438$.

Appendix E

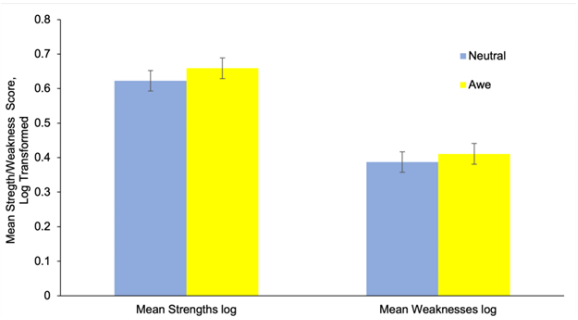

Figure 2. Average Log-Transformed Strength and Weakness Scores by Condition. The number of strengths and weaknesses that participants reported was positively skewed; therefore, we log transformed them. There was no significant difference between log-strength scores for the awe (M = 0.66, SD = 0.18) and neutral (M = 0.62, SD = 0.23) conditions $t(91) = 0.85$, $p = 0.397$. There was no significant difference between log-weakness scores for the awe (M = 0.41, SD = 0.24) and neutral (M = 0.39, SD = 0.19) conditions $t(82) = 0.50$, $p = 0.618$. Error bars show standard errors.

Appendix F

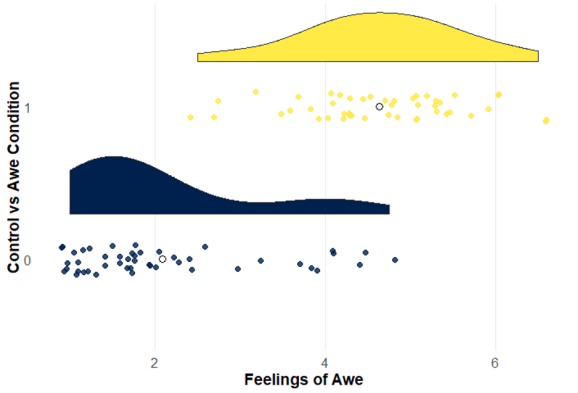

Figure 3. Differences in Reported Feelings of Awe Between Conditions. 0 = Neutral (control) condition participants in blue, 1 = awe condition participants in yellow. Results indicate that the experimental condition was a significant positive predictor of awe scores, $b_1 = 2.85$, $t(91) = 8.80$, $p < 0.001$. The average reported awe for participants in the neutral condition was 2.13, $t(91) = 9.33$, $p < 0.001$. Participant condition accounted for $R^2_{\text{adjusted}} = 45.4\%$ of the variance in reported awe.

Appendix G

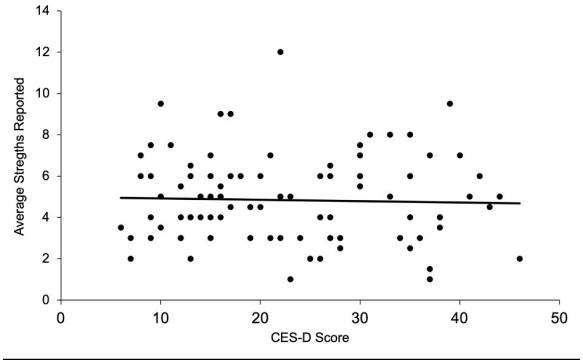

Figure 4. Changes in Reported Strengths as a Function of Depression Score. This figure shows the relationship between the number of participant-reported personal strengths, averaged between two raters, and their total score on the Center for Epidemiological Studies – Depression Scale (CES-D).