Perceptual & Motor Skills: Perception
2014, 119, 1, 237-253. © Perceptual & Motor Skills 2014

IDENTIFYING A RELIABLE BOREDOM INDUCTION[1, 2]

AMANDA MARKEY, ALYCIA CHIN, ERIC M. VANEPPS, AND GEORGE LOEWENSTEIN

Carnegie Mellon University

Summary.—None of the tasks used to induce boredom have undergone rigorous psychometric validation, which creates potential problems for operational equivalence, comparisons across studies, statistical power, and confounding results. This methodological concern was addressed by testing and comparing the effectiveness of six 5-min. computerized boredom inductions (peg turning, audio, video, signature matching, one-back, and an air traffic control task). The tasks were evaluated using standard criteria for emotion inductions: intensity and discreteness. Intensity, the amount of boredom elicited, was measured using a subset of the Multidimensional State Boredom Scale. Discreteness, the extent to which the task elicited boredom and did not elicit other emotions, was measured using a modification of the Differential Emotion Scale. In both a laboratory setting (Study 1; $N=241$) and an online setting with Amazon Mechanical Turk workers (Study 2; $N=416$), participants were randomly assigned to one of seven tasks (six boredom tasks or a comparison task, a clip from *Planet Earth*) before rating their boredom using the MSBS and other emotions using the modified DES. In both studies, each task had significantly higher intensity and discreteness than the comparison task, with moderate to large effect sizes. The peg-turning task outperformed the other tasks in both intensity and discreteness, making it the recommended induction. Identification of reliable and valid boredom inductions and systematic comparison of their relative results should help advance state boredom research.

A growing interest in state boredom has sparked an increased need for validated methods of boredom elicitation (for a review, see Eastwood, Frischen, Fenske, & Smilek, 2012). Boredom has been defined as an aversive and deactivated affective state in which individuals tend to experience a slow passage of time and a pervasive lack of interest, meaning, and engagement in current activities (e.g., Smith & Ellsworth, 1985; Nett, Goetz, & Daniels, 2010; Fahlman, Mercer-Lynn, Flora, & Eastwood, 2013; but see Mikulas & Vodanovich, 1993; Vogel-Walcutt, Fiorella, Carper, & Schatz, 2012). In contrast to the measurement of state boredom, which has received empirical attention (Van Tilburg & Igou, 2012; Fahlman, *et al.*, 2013), no researcher has compared the validity or relative effectiveness of tasks used to elicit boredom. This lack of validation can lead to underpowered and confounded studies, making it difficult to resolve conflicting findings (for a discussion of conflicting findings, see Merrifield, 2010, pp. 5–6). The cur-

[1]Address correspondence to Amanda Markey, Department of Social and Decision Sciences, Carnegie Mellon University, Baker Hall 321, 5000 Forbes Ave., Pittsburgh, PA 15213 or e-mail (amarkey@andrew.cmu.edu).
[2]This work was supported by the CBDR Small Grants program at Carnegie Mellon University. The authors thank Barbara Nash for her research assistance in collecting data.

DOI 10.2466/27.PMS.119c18z6 ISSN 0031-5125

rent research addressed this methodological gap in the validation of tasks commonly used to elicit boredom by comparing six boredom tasks to identify their relative effectiveness. The availability of validated boredom elicitations will allow researchers to conduct future studies with confidence.

Previous tasks for eliciting boredom can be classified into three broad categories: repetitive kinesthetic, simple cognitive, and media tasks. Repetitive kinesthetic tasks include repeatedly making check marks on paper (Geiwitz, 1966), writing "cd" (London, Schubert, & Washburn, 1972; Abramson & Stinson, 1977), screwing bolts and nuts together (Fisher, 1998), a hand-eye coordination task (Barmack, 1939), data entry (Lundberg, Melin, Evans, & Holmberg, 1993), copying references, tracing spirals, and connecting shapes by drawing lines between them (Van Tilburg & Igou, 2011). Existing simple cognitive tasks vary considerably. Researchers have used Tetris (Chanel, Rebetez, Bétrancourt, & Pun, 2008), proofreading address labels (Fisher, 1998), classifying whether objects are manmade (Jiang, Lianekhammy, Lawson, Guo, Lynam, Joseph, *et al.*, 2009), basic addition problems (Locke & Bryan, 1967), quantity approximation tasks (Van Tilburg & Igou, 2011), and counting tasks (Locke & Bryan, 1967; Van Tilburg & Igou, 2011). They also have used signal detection tasks, including monitoring a light on a box (London, *et al.*, 1972) and radar detection (Bailey, Thackray, Pearl, & Parish, 1976; Thackray, Bailey, Touchston, 1977; Hitchcock, Dember, Warm, Moroney, & See, 1999). Lastly, researchers have employed media tasks (audio and video clips) to elicit boredom. Rodin (1975) had participants listen to an excerpt from a textbook. Participants have also watched a video of men doing laundry (Merrifield & Danckert, 2014), a lesson of English as a Second Language, and a lecture on computer graphics (Fahlman, *et al.*, 2013).

The diversity of these boredom elicitations is a testament to researchers' creativity in designing experiments, but also reflects the complexity of and uncertainty in what causes boredom and the definition of what constitutes boredom. Differences in tasks make it difficult to resolve conflicting results when they arise. For example, researchers have found that boredom is associated with both an increase (London, *et al.*, 1972; Lundberg, *et al.*, 1993) and a decrease (Barmack, 1939) in heart rate. Unfortunately, these results were discovered using different inductions: a 30-min. writing task, 90 min. of slow-paced data entry, and a two-hour hand-eye coordination task, respectively. With such diverse tasks being used, and without rigorous task validation using standardized criteria, it is difficult to evaluate these studies and assess why their results conflict.

Psychometrics, the study of scale development and validation, has an extensive history and has standardized criteria for evaluating scales (c.f., Furr & Bacaharach, 2013), but efforts to validate emotion elicitations are

recent. The preferred method for evaluating emotion elicitations was established by Gross and Levenson (1995), who identified two criteria for selecting an emotion induction: *intensity* and *discreteness* (for discussions of these criteria, see Rottenberg, Ray, & Gross, 2007; Schaefer, Nils, Sanchez, & Philippot, 2010). Intensity refers to the amount of an emotion experienced by participants whereas discreteness refers to whether an experimental protocol elicits the target emotion without inducing other emotions. In the case of boredom, high intensity indicates that participants experienced boredom, whereas discreteness indicates that participants experienced boredom more than other emotions. Together, these criteria ensure that there is a strong manipulation of the target emotion—boredom—while reducing the confounding effects of other emotions.

In order to incorporate both intensity and discreteness into the selection of a task, Gross and Levenson (1995, p. 93; see also, Rottenberg, *et al.*, 2007, p. 18) advocated combining both metrics into a single measure, termed a Success Index. This index is calculated by normalizing intensity and discreteness across elicitations and then summing the two z-scores. Preferred elicitations have higher Success Indexes.

One secondary criterion (see Lang, Bradley, & Cuthbert, 1999, p. 2; Rottenberg, *et al.*, 2007, p. 18) for emotion inductions is *between-subjects reliability*, measured by the variance in the induced emotion. Higher between-subjects reliability, or lower variance, indicates a relatively uniform effect of boredom across individuals. Because higher reliability increases statistical power but does not affect internal validity, it is not included in the Success Index. However, it is useful for deciding between otherwise matched tasks.

Task Descriptions

Based on prior research in the area and a short pilot study in which participants rated anticipated boredom in different tasks,[3] six computerized boredom tasks and one comparison task were used. The six tasks span the different modal categories (i.e., repetitive kinesthetic, cognitive, and media). Each task lasted 5 min.

Peg turning.—Participants repeatedly clicked on icons of pegs that were arranged in two rows of four. Each mouse click rotated a peg a quarter turn clockwise, and participants were only able to click one peg at a time (it was highlighted). This repetitive kinesthetic task was adapted from the manual peg turning task used by Festinger and Carlsmith (1959) in their landmark study of cognitive dissonance.

[3]In an online pilot study ($N = 145$), participants rated how bored they would be if they engaged in each of several tasks for 15 minutes. The peg turning, 1-back, signatures, and air traffic control task were all included in this trial, and each elicited high ratings of anticipated boredom. The video and audio tasks described were developed and pre-tested separately.

1-back.—Participants viewed a series of digits (3 sec.) and fixation crosses (1 sec.). Digits were randomly selected with replacement from the set [0, 9]. Participants were instructed to press the spacebar when the digit that appeared was the same as the previous digit (33% of digits were the same as the previous digit).

Video.—Participants watched a video of a man talking about his work at an office supply company. He described, in a monotone and "boring" manner (Leary, Rogers, Canfield, & Coe, 1986), a conversation with a client, eating lunch at his desk, and the determinants of cardstock prices.

Audio.—Participants listened to the audio track of the video task, with no visual stimulus.

Air traffic control.—Participants saw a series of randomly ordered radar screens (2 sec), each of which showed two diagonal line segments representing airplanes. Radar screens either depicted the two planes on a collision course or on separate, non-colliding paths (3.3% of screens showed impending collisions, so participants saw approximately 5 impending collisions during the experiment; Hitchcock, *et al.*, 1999). Participants were instructed to press the spacebar when they saw an impending collision.

Signatures.—Participants viewed 20 pairs of signatures in a randomized order. After a forced waiting period of 15 sec., participants indicated whether the two signatures matched by clicking "yes" or "no" (signatures matched in 95% of cases). The next set of signatures appeared upon selection.

Planet Earth (comparison task).—Participants viewed a clip from *Mountains*, an episode of the British Broadcasting Company's (BBC) documentary film, *Planet Earth* (Fothergill, Berlowitz, Malone, & Lemire, 2007). This clip depicted nature and animal scenes, and was chosen because it has been shown to elicit interest and amusement without eliciting negative or positive emotions unrelated to boredom (e.g., disgust, relief; Bartolini, 2011; Merrifield & Danckert, 2014). This task was used to ensure that participants would not identify all tasks as boring, and thus provided a benchmark for comparison.

Current Study

This research compared six boredom tasks and one comparison task on their relative effectiveness in inducing boredom, as assessed by participants' ratings of the intensity and discreteness of their boredom. Two studies were conducted on different populations to gauge the generalizability of the results.

Study 1

Study 1 was designed to test the relative effectiveness of the six boredom inductions in a controlled laboratory setting. As a manipulation check, the intensity and discreteness ratings for each boredom task were compared to those from the comparison task, *Planet Earth*. All six bore-

dom tasks were predicted to be more intense and discrete than the comparison task. Next, intensity, discreteness, and between-subjects reliability were compared among the six boredom tasks. The Success Index was used to determine the most effective task. *A priori*, there was no hypothesis regarding which boredom task would have the highest Success Index.

Method

Participants

Participants were recruited for a laboratory experiment from the Carnegie Mellon Center for Behavioral Decision Research (CBDR) pool, a pool consisting of undergraduate and graduate students at Carnegie Mellon University and residents of Pittsburgh, Pennsylvania. Participants volunteered for a 10-min. "Attention Study" that was advertised on the CBDR website and were compensated with their choice of course credit or $3. Twenty-five participants (9.4%) were removed because they failed Task Fidelity screenings (see details below).[4] The final sample consisted of 241 participants (143 women, 184 Caucasian, M age = 31.9 yr., SD = 13.0).[5]

Measures

State boredom.—Participants indicated their agreement (1:Strongly disagree; 7:Strongly agree) with seven statements from a validated measure of state boredom, the Multidimensional State Boredom Scale (MSBS; Fahlman, *et al.*, 2013).[6] The seven statements were: "Time was passing by slower than usual"; "I was stuck in a situation that I felt was irrelevant"; "Everything seemed repetitive and routine to me"; "I felt bored"; "I seemed to be forced to do things that have no value to me"; "I wished I were doing something more exciting." Responses were averaged to create a boredom rating (Cronbach's αs = .90 and .92 in Studies 1 and 2). Intensity was measured as the mean boredom rating using this seven-item subset of the MSBS, and between-subjects reliability was measured as the variance of these ratings.

---

[4]Ten participants had an average Task Fidelity Rating of 4.0 or below; an additional 15 indicated that their data should not be used.

[5]In the final data set, age was negatively correlated to boredom intensity, $r = -.26$, $p < .0005$. There was no significant difference in intensity between sexes ($p = .14$) or between Caucasian and non-Caucasian participants ($p = .60$).

[6]The Multidimensional State Boredom Scale (MSBS; Fahlman, 2008; Fahlman, *et al.*, 2013) was developed across multiple experiments using a total of nearly 2,000 participants to ensure strong psychometric properties. The scale exhibited high reliability ($\alpha = .94$). It possessed high convergent validity, as established by its moderate to high correlations with two existing measures of trait boredom and with related constructs, including depression, anxiety, anger, inattention, neuroticism, impulsivity, decreased purpose in life, and decreased life satisfaction. Finally, the authors established the predictive validity of the scale by showing participants either an interesting or boring video clip and then measuring their level of boredom using an open-ended response and the MSBS. As expected, participants in the boring condition described the video clip as "boring" in their open-ended responses and also scored higher on the MSBS. For more details, see Fahlman (2008).

Differential emotions.—The emotion measure was a modified version of the Differential Emotion Scale (DES; Gross & Levenson, 1995) that included a 17-item emotion self-report inventory (i.e., amusement, anger, arousal, boredom, confusion, contempt, contentment, disgust, embarrassment, fear, happiness, interest, pain, relief, sadness, surprise, tension). Participants indicated the greatest amount of each emotion they felt during the task on a 9-point scale with anchors 0: Not even the slightest bit and 8: The most I've ever felt in my life. Following the procedures of Gross and Levenson (1995; Rottenberg, *et al.*, 2007, p. 18), discreteness was measured as the percentage of participants whose boredom rating was at least one point higher than the other main emotion terms on the modified DES.[7]

Success Index.—The Success Index (Gross & Levenson, 1995, pp. 92–93; Rottenberg, *et al.*, 2007, p. 18) was calculated as the sum of the normalized intensity and discreteness measures across the six tasks.

Task fidelity.—Participants indicated their agreement with three statements using a seven-point Likert-type scale (1: Strongly disagree; 7: Strongly agree). An example item is, "I tried my hardest to do this task as instructed." Cronbach's αs were .70 and .69 in Studies 1 and 2, respectively. Data from respondents who had an average Task Fidelity Rating of 4 or below ("Neutral") were dropped. Additionally, participants were dropped who said "no" to the question, "In your honest opinion, should we use your data?" (Meade & Craig, 2012). Finally, in the Amazon Mechanical Turk sample (Study 2), observations with duplicate IP addresses were eliminated. All elimination criteria were determined prior to data collection.

Procedure

In a between-subjects design, participants were randomly assigned to one of the seven tasks (six elicitations plus one comparison task) and subsequently answered questions about their State Boredom, their Differential Emotional reactions, and their Task Fidelity. The study ended with open-ended comments and demographic questions (age, sex, and ethnicity).

Results

Comparisons (Mann-Whitney U) showed that all six boredom tasks induced more intense boredom than *Planet Earth* ($U > 57$, $p < .0005$), with large effect sizes (r between .51 and .77; Table 1).[8] Furthermore, those who

[7]This study adopts Gross and Levenson's (1995) calculation of discreteness, which uses seven emotion terms: amusement, anger, contentment, disgust, fear, sadness, and surprise. When discreteness is calculated using all 17 emotion terms, the same pattern of results holds.

[8]The effect size of the Mann-Whitney test, r, ranges from 0 to 1 and is calculated by dividing the z statistic by the square root of the sample size (for reference, see Corder & Foreman, 2009, pp. 39–40). Cohen (1988; 1992) established conventions for interpreting effect size, with r considered to be small (.10), medium (.30), or large (.50).

TABLE 1
SUMMARY STATISTICS BY TASK, STUDY 1

Task	Intensity			Discreteness		Success Index	*n*
	M	*SD*	Effect Size (*r*)	%	Effect Size (φ)		
1. Peg turning	5.59[a, b]	1.02	0.77	61.8[a]	0.544	1.78	34
2. Video	4.90[a]	1.13	0.68	71.9[a]	0.636	1.70	32
3. Audio	4.82[a]	0.87	0.67	68.8[a]	0.608	1.05	32
4. 1-back	4.53[a]	1.29	0.55	57.9[a]	0.504	−1.22	38
5. Air traffic control	4.44[a]	1.37	0.51	56.8[a]	0.496	−1.59	37
6. Signatures	4.46[a]	1.10	0.61	55.6[a]	0.487	−1.72	36
7. *Planet Earth* (comparison task)	2.91	1.17	—	9.4	—	—	32
Mean, All Tasks	*4.78*	*1.21*	—	*61.2*	—	—	*241*

Note.—Boredom was a composite measure with a possible range of 1 to 7. Higher scores indicate more intense boredom. Discreteness was equal to the percentage of participants whose boredom rating on the Differential Emotion Scale was at least one point higher than the other discrete emotion terms: amusement, anger, contentment, disgust, fear, sadness, and surprise. The Success Index was calculated by normalizing intensity and discreteness for all of the tasks and summing the two *z*-scores. The effect size *r* reflects pairwise comparisons to *Planet Earth* and was computed as the Mann Whitney *z* statistic divided by the square root of the sample size. [a]*Planet Earth* comparison, $p < .0005$. [b]Average of other five boring tasks comparison, $p < .05$.

completed a boredom task were more likely to experience boredom as a discrete emotion than those who watched *Planet Earth* [χ^2 (1)>16.14, $p < .0005$], again with large effect sizes (φ between 0.48 and 0.64; Table 1).[9,10]

A Kruskal-Wallis nonparametric test indicated that there were significant differences among the tasks in boredom intensity [H(5)=22.53, $p < .0005$]. Follow-up comparisons indicated that the peg-turning task was significantly more boring than the average of the other five boredom tasks [$U = 1597$, $p < .0005$, $r = .30$].[11] Discreteness was not significantly different among the six boredom tasks [$\chi^2(5) = 3.27$, $p = .66$]. Between-subjects reliability was not significantly different among the six boredom tasks (Brown-Forsythe $F_{5,203} = 2.11$, $p = .07$]. Overall, the Success Index was highest for the peg-turning task, followed by the video.

[9]The effect size of the chi-squared test, phi (ϕ), is calculated as the square root of the chi-squared test statistic divided by the sample size. Like *r*, the statistic ranges from 0 to 1, and follows Cohen's (1988; 1992) conventions of .10, .30, and .50 to indicate small, medium or large effects.
[10]The comparison task, *Planet Earth*, had generally higher single-item ratings of amusement, happiness, and interest than the six boredom tasks, consistent with the definition of boredom and with past research using this stimulus (see Appendix, Table A).
[11]Additionally, exploratory analyses revealed that every one of the seven items used to measure boredom was highest for the peg turning task (see Appendix, Table B).

Discussion

Study 1 indicated that all six of the experimental tasks were rated more boring and more discrete than the comparison task, *Planet Earth*. Furthermore, Study 1 provided evidence that the peg-turning task elicited significantly more boredom intensity than any of the other boredom tasks, and the Success Index was highest for the peg-turning task.

Study 1 was constrained to local residents and university students who were recruited from a research participation pool. Study 2 used identical procedures and measures with participants from a different population to explore whether these patterns generalized.

Study 2

To test the relative effectiveness of the six boredom inductions in a different experimental setting, participants were recruited from Amazon's Mechanical Turk, an online marketplace for computerized work (Buhrmester, Kwang, & Gosling, 2011). As in Study 1, the intensity and discreteness of each boredom task were first compared to the comparison task, *Planet Earth*. Then intensity, discreteness, and between-subjects reliability were compared among the six boredom tasks. The Success Index was used to identify the most effective induction.

Participants

Participants were recruited using Amazon Mechanical Turk (www.mturk.com; Buhrmester, *et al.*, 2011). Participants volunteered for a 10 min. "Attention Study" that was advertised on the website. Eligible participants had at least a 95% approval rating on previous tasks. Participants who failed Task Fidelity screenings were excluded from analysis ($N = 44$).[12] The final sample consisted of 416 participants (166 women; 243 Caucasian; M age = 33.3 yr., $SD = 11.3$).[13]

Results and Discussion

All six boredom tasks were rated significantly more boring than the comparison task (Mann-Whitney Us > 326, $p < .0005$), with moderate to large effect sizes for all tasks (r between .39 and .66; Table 2). Additionally, those who completed any of the boredom tasks were more likely to rate boredom as a discrete emotion than those who watched the *Planet Earth* video [$\chi^2(1) > 13.02$, $p < .0005$], also exhibiting moderate to large effect sizes (φ between 0.32 and 0.56; Table 2).[14]

---

[12]Sixteen participants had an average Task Fidelity Rating 4.0 or below, another 7 participants indicated that their data should not be used, and an additional 21 responses from duplicate IP addresses were dropped.

[13]Age was negatively correlated to boredom ($r = -.35$, $p = .01$). There was no significant difference across sex ($p = .45$) or between Caucasian and non-Caucasian participants ($p = .11$).

[14]As in Study 1, the comparison task, *Planet Earth*, was generally higher in single-item ratings of amusement and interest compared to the six boredom tasks (see Appendix, Table C).

TABLE 2
SUMMARY STATISTICS BY TASK, STUDY 2

Task	Intensity			Discreteness		Success Index	*n*
	M	*SD*	Effect Size (*r*)	%	Effect Size (φ)		
1. Peg turning	5.79[a,b]	1.07	0.66	72.9[a]	0.56	3.09	48
2. 1-back	5.01[a]	1.19	0.49	67.8[a]	0.51	0.50	59
3. Air traffic control	5.15[a]	1.18	0.53	64.2[a]	0.47	0.48	67
4. Video	4.98[a]	1.18	0.48	61.5[a]	0.45	−0.26	52
5. Audio	4.98[a]	1.44	0.47	52.2[a]	0.36	−1.26	67
6. Signatures	4.64[a]	1.22	0.39	48.5[a]	0.33	−2.55	66
7. *Planet Earth* (comparison task)	3.37	1.67	—	17.5	—	—	57
Mean, All Boring Tasks	*5.06*	*1.29*	—	*60.5*	—	—	*416*

Note.—Boredom was a composite measure with a possible range of 1 to 7. Higher scores indicate more intense boredom. Discreteness was equal to the percentage of subjects whose boredom rating on the Differential Emotion Scale was at least one point higher than the other discrete emotion terms: amusement, anger, contentment, disgust, fear, sadness, and surprise. The Success Index was calculated by normalizing intensity and discreteness for all of the tasks and summing the two z-scores. The effect size *r* reflects pairwise comparisons to *Planet Earth* and was computed as the Mann Whitney z statistic divided by the square root of the sample size. [a]*Planet Earth* comparison, $p < .0005$. [b]Average of other five boring tasks comparison, $p < .05$.

A Kruskal-Wallis nonparametric test indicated significant differences among the six tasks in boredom intensity [$H(5) = 27.45$, $p < .0005$]. Peg turning was rated significantly more boring than the average of the other five boredom tasks ($U = 4410.5$, $p < .0005$, $r = .25$).[15] Discreteness was not different among the six boredom tasks [$\chi^2(5) = 10.71$, $p = .06$]. Between-subjects reliability among the six boring tasks was significantly different (Brown-Forsythe $F_{6, 409} = 3.32$, $p = .003$). However, no single task had a significantly lower variance than the combined variance of the other five boredom tasks ($p > .11$). The Success Index was highest for the peg-turning task, followed by the 1-back.

GENERAL DISCUSSION

This research is the first to systematically compare boredom inductions in terms of their overall effectiveness. All six computerized tasks tested here elicited more boredom than the comparison task, *Planet Earth*, with effect sizes ranging from 0.39 to 0.77. Additionally, all six tasks induced boredom without eliciting other emotions, with effect sizes of discreteness ranging from 0.33 to 0.64. Overall, the peg-turning task had the

[15]As in Study 1, exploratory analyses revealed that every one of the seven items used to measure boredom was highest for the peg turning task (see Appendix, Table D).

highest Success Index in both samples among the boredom tasks studied, making it the recommended induction. It is impressive that a repetitive motor task designed over 50 years ago by Festinger and Carlsmith emerged as the most boring task in the present day.

The data presented here provide a baseline task that can be used to explore boredom's causes and to test interventions. With only minor revisions to the peg-turning task, researchers could evaluate the effect of goals, performance feedback, cognitive load, perceptual stimuli (i.e., background music), task duration, and other characteristics on boredom. Manipulations such as these have the potential to inform theory and also practical interventions that minimize boredom in everyday life.

This research also allows researchers to make informed decisions regarding the selection of tasks to fit their study design. For instance, one may prefer to use videos to induce emotions in online participants. The effect sizes, sample sizes, and task details reported here suggest that a comparison of the boredom video and the *Planet Earth* video would require approximately 38 laboratory participants to have 0.80 power to detect a significant difference in boredom between conditions, using a two-tailed test and setting a conservative α of 0.001 (Cohen, 1988, 1992).

Limitations and Future Directions

All of the tasks in this work were computerized, as one goal of this research was to create flexible stimuli that can be easily adapted by future researchers. However, boredom can emerge in situations that are not reflected in a computerized stimulus. Although the present tasks span the range of boredom inductions used in previous research (i.e. repetitive kinesthetic, cognitive, and media), they still are only an arbitrary selection of possible tasks in daily life. Researchers should be mindful of circumstances in which alternative inductions provide improved external validity.

Second, the evaluation of the boredom tasks depends on the scale used to measure boredom. In this study, boredom was measured with a subset of the MSBS, a measure that includes questions on perceived meaning, repetition, and slowed time. Although most boredom researchers would agree that these characteristics are associated with boredom (for a discussion, see Mikulas & Vodanovich, 1993), adopting a different definition of boredom would result in a different evaluation of the tasks. This concern is partially alleviated because the peg-turning task scored the highest on every item in the seven-item MSBS subset (most notably, "I felt bored"), and it also scored the highest on the single-item measure using the modified DES. However, the lack of a consensus regarding the definition of boredom remains a fundamental obstacle facing the field, and necessarily limits the applicability of any single induction.

Finally, the present work does not explicitly test why these tasks induce boredom or why they are differentially effective. Although this study did not predict which task would be most effective at eliciting boredom, in hindsight one can speculate on why the peg-turning task was the most successful. First, the task is highly repetitive, as participants not only repeatedly click pegs but are also forced to "start over" after every set of pegs. The association between repetition and boredom has been discussed since the early 1900's in the context of assembly line work (Wyatt, 1929), and turning pegs exemplifies such repetition. Second, peg turning may be particularly boring because it is not challenging. In his work on boredom and flow, Csikszentmihalyi (2000) suggested that boredom occurs when skill exceeds challenge, i.e., when tasks are too easy. In contrast to the other tasks examined in this study, peg turning requires minimal cognitive effort. Finally, peg turning may feel meaningless. Whereas the air traffic control task may have real world parallels that lend it a sense of importance, it is difficult to imagine the usefulness of repeatedly clicking a button to rotate a circle on a screen. Although the current data are limited in their ability to specify what makes each task more or less boring, future work that manipulates task characteristics, ranging from repetition to cognitive load to active or passive engagement, may allow researchers to disentangle the multiple elements that drive boredom.

By creating, validating, and demonstrating the effectiveness of six different boredom inductions, the current work provides an important methodological foundation for future researchers interested in studying boredom across a broad range of experimental paradigms. Having validated methods for inducing boredom, assessing their relative effect sizes, and encouraging the application of the new state boredom scale (Fahlman, *et al.*, 2013) should facilitate continued research on a wide variety of judgments, choices, and behaviors that may be affected by boredom.

REFERENCES

ABRAMSON, E. E., & STINSON, S. G. (1977) Boredom and eating in obese and non-obese individuals. *Addictive Behaviors*, 2(4), 181-185.

BAILEY, J. P., THACKRAY, R. I., PEARL, J., & PARISH, T. S. (1976) Boredom and arousal: comparison of tasks differing in visual complexity. *Perceptual & Motor Skills*, 43(1), 141-142.

BARMACK, J. E. (1939) Studies on the psychophysiology of boredom: Part I. The effect of 15 mgs. of benzedrine sulfate and 60 mgs. of ephedrine hydrochloride on blood pressure, report of boredom and other factors. *Journal of Experimental Psychology*, 25(5), 494.

BARTOLINI, E. E. (2011) Eliciting emotion with film: development of a stimulus set. Unpublished doctoral dissertation, Wesleyan Univer.

BUHRMESTER, M., KWANG, T., & GOSLING, S. (2011) Amazon's Mechanical Turk: a new source of inexpensive, yet high-quality, data? *Perspectives on Psychological Science*, 6, 3-5.

CHANEL, G., REBETEZ, C., BÉTRANCOURT, M., & PUN, T. (2008) Boredom, engagement and anxiety as indicators for adaptation to difficulty in games. *Proceedings of the 12th International Conference on Entertainment and Media in the Ubiquitous Era.* New York: ACM. Pp. 13-17.

COHEN, J. (1988) *Statistical power analysis for the behavioral sciences.* (2nd ed.) Hillsdale, NJ: Erlbaum.

COHEN, J. (1992) A power primer. *Psychological Bulletin*, 112(1), 155-159.

CORDER, G. W., & FOREMAN, D. I. (2009) *Nonparametric statistics for non-statisticians: a step-by-step approach.* Hoboken, NJ: Wiley.

CSIKSZENTMIHALYI, M. (2000) *Beyond boredom and anxiety.* San Francisco, CA: Jossey-Bass.

EASTWOOD, J. D., FRISCHEN, A., FENSKE, M. J., & SMILEK, D. (2012) The unengaged mind: defining boredom in terms of attention. *Perspectives on Psychological Science*, 7, 482-495.

FAHLMAN, S. (2008) Development and validation of the Multidimensional State Boredom Scale. Unpublished doctoral thesis, York University, Toronto, ON.

FAHLMAN, S. A., MERCER-LYNN, K. B., FLORA, D. B., & EASTWOOD, J. D. (2013) Development and validation of the Multidimensional State Boredom Scale. *Assessment*, 20, 68-85.

FESTINGER, L., & CARLSMITH, J. M. (1959) Cognitive consequences of forced compliance. *Journal of Abnormal and Social Psychology*, 58, 203-210.

FISHER, C. D. (1998) Effects of external and internal interruptions on boredom at work: two studies. *Journal of Organizational Behavior*, 19(5), 503-522.

FOTHERGILL, A. (DIRECTOR), BERLOWITZ, V., MALONE, S., & LEMIRE, M. (PRODUCERS) (2007) *Planet Earth.* [Film] Burbank, CA: Warner Home Video.

FURR, R. M., & BACHARACH, V. R. (2013) *Psychometrics: an introduction.* London, UK: Sage.

GEIWITZ, P. J. (1966) Structure of boredom. *Journal of Personality and Social Psychology*, 3(5), 592-600.

GROSS, J. J., & LEVENSON, R. W. (1995) Emotion elicitation using films. *Cognition & Emotion*, 9(1), 87-108.

HITCHCOCK, E. M., DEMBER, W. N., WARM, J. S., MORONEY, B. W., & SEE, J. E. (1999) Effects of cueing and knowledge of results on workload and boredom in sustained attention. *Human Factors*, 41, 365-372.

JIANG, Y., LIANEKHAMMY, J., LAWSON, A., GUO, C., LYNAM, D., JOSEPH, J. E., GOLD, B. T., & KELLY, T. H. (2009) Brain responses to repeated visual experience among low and high sensation seekers: role of boredom susceptibility. *Psychiatry Research: Neuroimaging*, 173(2), 100-106.

LANG, P. J., BRADLEY, M. M., & CUTHBERT, B. N. (1999) *International Affective Picture System (IAPS): instruction manual and affective ratings.* Gainesville, FL: The Center for Research in Psychophysiology, Univer. of Florida.

LEARY, M. R., ROGERS, P. A., CANFIELD, R. W., & COE, C. (1986) Boredom in interpersonal encounters: antecedents and social implications. *Journal of Personality and Social Psychology*, 51(5), 968-975.

LOCKE, E. A., & BRYAN, J. F. (1967) Performance goals as determinants of level of performance and boredom. *Journal of Applied Psychology*, 51(2), 120-130.

LONDON, H., SCHUBERT, D. S., & WASHBURN, D. (1972) Increase of autonomic arousal by boredom. *Journal of Abnormal Psychology*, 80(1), 29-36.

LUNDBERG, U., MELIN, B., EVANS, G. W., & HOLMBERG, L. (1993) Physiological deactivation after two contrasting tasks at a video display terminal: learning vs repetitive data entry. *Ergonomics*, 36(6), 601-611.

MEADE, A. S., & CRAIG, S. B. (2012) Identifying careless responses in survey data. *Psychological Methods*, 17, 437-455.

MERRIFIELD, C. (2010) Characterizing the psychophysiological signature of boredom. Unpublished master's thesis, Univer. of Waterloo, Waterloo, Ontario.

MERRIFIELD, C., & DANCKERT, J. (2014) Characterizing the psychophysiological signature of boredom. *Experimental Brain Research*, 232(2), 481-491.

MIKULAS, W. L., & VODANOVICH, S. J. (1993) The essence of boredom. *The Psychological Record*, 43(1), 3-12.

NETT, U. E., GOETZ, T., & DANIELS, L. M. (2010) What to do when feeling bored? Students' strategies for coping with boredom. *Learning and Individual Differences*, 20(6), 626-638.

RODIN, J. (1975) Causes and consequences of time perception differences in overweight and normal weight people. *Journal of Personality and Social Psychology*, 31(5), 898-904.

ROTTENBERG, J., RAY, R., & GROSS, J. (2007) Emotion elicitation using films. In J. A. Coan & J. J. B. Allen (Eds.), *The handbook of emotion elicitation and assessment.* New York: Oxford Univer. Press.

SCHAEFER, A., NILS, F., SANCHEZ, X., & PHILIPPOT, P. (2010) Assessing the effectiveness of a large database of emotion-eliciting films: a new tool for emotion researchers. *Cognition & Emotion*, 24(7), 1153-1172.

SMITH, C. A., & ELLSWORTH, P. C. (1985) Patterns of cognitive appraisal in emotion. *Journal of Personality and Social Psychology*, 48(4), 813-838.

THACKRAY, R. I., BAILEY, J. P., & TOUCHSTONE, R. M. (1977) Physiological, subjective, and performance correlates of reported boredom and monotony while performing a simulated radar control task. In R. R. Mackie (Ed.), *Vigilance: theory, operational performance and physiological correlates.* New York: Plenum. Pp. 203-216.

VAN TILBURG, W. A. P., & IGOU, E. R. (2011) On boredom and social identity: a pragmatic meaning-regulation approach. *Personality and Social Psychology Bulletin*, 37, 1679-1692.

VOGEL-WALCUTT, J. J., FIORELLA, L., CARPER, T., & SCHATZ, S. (2012) The definition, assessment, and mitigation of state boredom within educational settings: a comprehensive review. *Educational Psychology Review*, 24(1), 89-111.

WYATT, S. (1929) Boredom in industry. *Personnel Journal*, 8(3), 161-171.

Accepted July 2, 2014.

APPENDIX

TABLE A

DIFFERENTIAL EMOTION SCALE IN LABORATORY SETTING, STUDY 1

Task	Peg Turning		Video		Audio		1-back		Signatures		Air Traffic Control		*Planet Earth*	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Amusement	2.44*	2.29	1.63*	1.56	2.13*	2.14	2.24*	1.91	2.47*	2.08	2.49*	2.01	4.34	1.89
Anger	0.97	1.90	0.72	1.63	0.56	1.08	0.42	1.08	0.78	1.59	0.41	1.07	0.34	0.87
Arousal	0.35	0.98	0.34	0.65	0.47	1.39	0.79	1.47	0.64	1.52	0.32	0.63	1.00	1.48
Boredom	4.59*	2.49	4.66*	2.28	4.13*	1.88	4.05*	2.39	4.06*	2.19	3.62*	1.75	1.91	2.29
Confusion	1.68	2.16	2.09	2.13	1.72	1.69	1.21	1.88	2.19	1.98	0.97	1.61	1.03	1.81
Contempt	0.97	1.75	0.53	1.02	1.00	1.50	0.76	1.52	0.64	1.13	0.38	0.86	0.91	2.07
Contentment	1.24	1.56	2.03	1.64	1.81	1.77	1.74	1.83	1.78	1.71	2.24	2.14	3.63	2.34
Disgust	0.68	1.65	0.63	1.36	0.31	0.86	0.63	1.34	0.67	1.45	0.03	0.16	0.22	0.75
Embarrassment	0.56	1.35	0.31	0.90	0.38	0.94	0.53	0.92	0.36	1.07	0.24	0.64	0.25	1.41
Fear	0.47	1.33	0.06	0.25	0.06	0.25	0.39	1.10	0.25	0.65	0.68	1.51	0.16	0.63
Happiness	1.44*	1.78	1.03*	1.20	1.13*	1.26	1.61*	1.84	1.72*	1.99	1.78*	2.07	3.69	2.10
Interest	2.12*	1.97	2.16*	1.48	2.25*	1.92	3.32	2.34	2.89*	1.98	2.70*	2.00	5.19	2.22
Pain	0.50	1.08	0.38	1.07	0.13	0.55	0.24	0.79	0.28	0.88	0.24	0.72	0.38	1.21
Relief	2.00	2.85	0.59	1.04	1.50	2.02	1.11	1.71	1.33	1.85	1.19	1.82	0.94	1.78
Sadness	0.12	0.41	0.88	1.79	0.44	0.88	0.13	0.48	0.58	1.44	0.19	0.62	0.25	0.67
Surprise	1.21	2.25	0.59*	1.10	0.69*	1.45	0.97	1.64	1.03	1.54	1.16	1.57	2.34	2.31
Tension	2.03	2.36	1.47	1.83	1.34	1.70	1.76	2.05	1.11	1.70	1.54	1.73	0.69	1.18
Discreteness	0.62		0.72		0.69		0.58		0.56		0.57		0.09	
n	34		32		32		38		36		37		32	

Note.—Each emotion was rated on a scale with anchors 0: Not even the slightest bit and 8: The most I've ever felt in my life. **Planet Earth* comparison, $p < .0005$.

TABLE B

MULTIDIMENSIONAL BOREDOM SCALE ITEMS IN LABORATORY SETTING, STUDY 1

Task	Item														*n*
	1		2		3		4		5		6		7		
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	
Peg turning	5.44	1.40	5.68	1.25	5.59	1.37	6.38	0.95	5.76	1.37	4.76	1.71	5.53	1.52	34
Video	4.91	1.28	5.25	1.37	4.91	1.51	4.66	1.41	5.00	1.46	4.50	1.55	5.06	1.46	32
Audio	4.69	1.20	5.13	1.13	5.09	1.33	4.31	1.26	5.34	1.10	4.25	1.22	4.91	1.45	32
1-back	4.37	1.65	3.89	1.91	5.26	1.45	4.97	1.55	4.79	1.79	3.47	1.86	5.00	1.86	38
Air traffic control	4.41	1.46	3.59	1.82	4.68	1.77	5.08	1.57	5.05	1.79	3.65	2.00	4.59	1.82	37
Signatures	5.00	1.12	3.75	1.68	4.56	1.58	4.67	1.39	4.75	1.54	3.78	1.82	4.72	1.68	36
Planet Earth (comparison task)	3.63	1.29	2.75	1.22	3.06	1.50	2.56	1.41	2.75	1.74	2.69	1.51	2.94	1.65	32

Note.—Items were: 1. Time was passing by slower than usual. 2. I was stuck in a situation that I felt was irrelevant. 3. I wished time would go by faster. 4. Everything seemed repetitive and routine to me. 5. I felt bored. 6. I seemed to be forced to do things that have no value to me. 7. I wished I were doing something more exciting. Respondents indicated their agreement using a seven-point Likert-type scale (1: Strongly disagree; 2: Disagree; 3: Somewhat disagree; 4: Neutral; 5: Somewhat agree; 6: Agree; and 7: Strongly agree).

TABLE C
DIFFERENTIAL EMOTION SCALE IN ONLINE SETTING, STUDY 2

Emotion	Peg Turning		Video		Audio		1-back		Signatures		Air Traffic Control		*Planet Earth*	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Amusement	2.06*	1.98	2.04*	2.22	1.90*	1.93	2.46*	2.36	2.86*	2.29	2.31*	2.26	4.44	2.01
Anger	1.77*	2.13	1.38	2.05	1.30	2.13	1.31	2.14	0.68	1.17	0.85	1.51	0.44	0.96
Arousal	1.46	2.07	1.04	1.99	1.13	1.89	1.05	1.80	1.00	1.63	0.82	1.39	1.18	1.62
Boredom	5.33*	2.50	4.63*	2.38	4.55*	2.61	4.61*	2.24	3.89	2.09	4.88*	2.43	2.56	2.38
Confusion	2.38	2.28	1.77	1.98	1.97	2.44	1.58	2.20	1.92	2.14	1.45	2.21	1.18	1.79
Contempt	1.98*	2.26	1.27	1.96	1.37	2.13	1.64	1.98	1.17	1.76	1.15	1.86	0.61	1.26
Contentment	2.13	2.10	1.83	1.88	1.85*	2.04	1.59	1.75	2.38	2.06	1.97	2.10	3.07	2.23
Disgust	1.46	2.21	0.94	1.84	1.30	2.10	1.03	1.83	0.59	1.25	0.91	1.73	0.54	1.36
Embarrassment	1.25	2.08	0.77	1.75	0.94	1.94	0.86	1.91	0.56	1.28	0.69	1.35	0.58	1.45
Fear	0.52	1.41	0.40	1.02	0.78	1.86	0.61	1.57	0.53	1.26	0.58	1.47	0.32	0.87
Happiness	2.17	2.43	2.17	2.09	2.19*	2.40	2.07	2.17	2.53	2.34	2.51	2.54	3.68	2.59
Interest	2.42*	2.31	2.79*	2.34	2.96*	2.47	2.97*	2.38	3.82	2.20	3.40*	2.56	5.19	2.13
Pain	1.19	2.08	0.62	1.22	0.96	1.88	0.81	1.91	0.42	1.10	0.55	1.31	0.33	0.93
Relief	1.52	2.05	1.87	2.21	2.03	2.47	1.61	2.13	1.86	2.08	2.00	2.11	1.23	1.89
Sadness	1.13	1.88	0.75	1.38	0.96	1.96	0.68	1.58	0.67	1.27	0.64	1.56	0.46	1.02
Surprise	1.65	2.36	1.25	1.84	1.12*	1.97	1.51	2.28	1.15*	1.68	1.63	2.46	2.58	2.36
Tension	2.06*	1.98	1.65	2.10	1.37	2.25	2.39*	2.54	1.39	1.78	1.94	2.38	0.81	1.70
Discreteness	0.73		0.63		0.52		0.68		0.49		0.64		0.18	
n	48		52		67		59		66		67		57	

Note.—Each emotion was rated on a scale with anchors 0: Not even the slightest bit and 8: The most I've ever felt in my life. **Planet Earth* comparison, $p < .0005$.

TABLE D

MULTIDIMENSIONAL BOREDOM SCALE ITEMS IN ONLINE SETTING, STUDY 2

Task	Item														n
	1		2		3		4		5		6		7		
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	
Peg turning	5.58	1.33	5.71	1.68	6.06	1.04	6.52	0.74	5.67	1.89	5.38	1.62	5.63	1.79	48
1-back	5.02	1.48	4.36	1.63	5.25	1.59	5.36	1.30	5.07	1.72	4.49	1.81	5.54	1.19	59
Air traffic control	5.25	1.35	4.49	1.71	5.63	1.35	5.69	1.34	5.34	1.60	4.39	1.78	5.27	1.61	67
Video	5.10	1.32	5.00	1.48	5.37	1.31	4.46	1.46	5.19	1.69	4.65	1.62	5.12	1.40	52
Audio	5.07	1.54	5.09	1.63	5.18	1.68	4.69	1.63	4.87	2.01	4.67	1.80	5.31	1.61	67
Signatures	5.00	1.56	3.94	1.52	5.20	1.49	4.92	1.60	4.61	1.75	3.94	1.74	4.88	1.45	66
Planet Earth (comparison task)	3.68	1.85	3.25	1.73	3.68	1.95	2.93	1.59	3.11	1.90	3.21	1.79	3.70	2.01	57

Note.—Items were: 1. Time was passing by slower than usual. 2. I was stuck in a situation that I felt was irrelevant. 3. I wished time would go by faster. 4. Everything seemed repetitive and routine to me. 5. I felt bored. 6. I seemed to be forced to do things that have no value to me. 7. I wished I were doing something more exciting. Respondents indicated their agreement using a seven-point Likert-type scale (1: Strongly disagree; 2: Disagree; 3: Somewhat disagree; 4: Neutral; 5: Somewhat agree; 6: Agree; and 7: Strongly agree).