American Psychological Association

Journal of Personality and Social Psychology:
Personality Processes and Individual Differences

0022-3514/19/$12.00

2019, Vol. 117, No. 2, e27–e34
http://dx.doi.org/10.1037/pspp0000199

Thinking About a New Decade in Life Increases Personal Self-Reflection: A Replication and Reinterpretation of Alter and Hershfield's (2014) Findings

Jinhyung Kim and Rebecca J. Schlegel
Texas A&M University

Elizabeth Seto
Colby College

Joshua A. Hicks
Texas A&M University

Alter and Hershfield (2014) recently published a set of studies suggesting that people often search for existential meaning as they approach a new decade in chronological age. The purpose of the current research was to replicate their experimental study (Study 2 in their article) and extend their findings using additional operational measures of search for meaning. Study 1 was a replication comparing the two conditions used in the original study (i.e., experimental and baseline control), whereas Studies 2 and 3 were direct replications of the original methods using all three conditions (i.e., experimental, baseline control, and birthday control). All replications found general support for the original claims with important caveats. Specifically, whereas Studies 1 and 3 replicated their main findings, Study 2 did not. Importantly, however, a factor analysis of Alter and Hershfield's meaning-seeking measure revealed two factors underlying a search for meaning: life-reflection and perceived value of meaning. Across all studies, findings suggest that people are significantly more likely to engage in a life review as they begin a new epoch in their lives while there were no differences in their perceived value of meaning. A reinterpretation of Alter and Hershfield's findings is discussed.

Keywords: age, life reflection, replication, search for meaning, time

Alter and Hershfield (2014) investigated the idea that people are more inclined to search for meaning as they approach a new decade in chronological age (e.g., before turning 30, 40, 50, etc.). Across six studies, they showed that participants who were approaching (or imagining approaching) a new decade in their life explicitly reported a preoccupation with existential concerns such as meaning and purpose (Studies 1 and 2). They were also more likely to engage in various behaviors that may reflect searching for meaning, such as seeking an extramarital affair or running their first marathon (Studies 3–6). The paper was published in a highly prestigious outlet, *The Proceedings of the National Academy of Sciences of the United States of America* (PNAS), and received prominent press coverage (36 news outlets according to Altmetric, including *ABC News, Time*, and *Reader's Digest*). The overall "attention score" by Altmetric placed it in the top 5% of all research scored by Altmetric.

Most of the studies used established available data (e.g., World Values Survey, CDC data on suicide). However, Study 2 reports an experimental study ripe for replication. The present article reports our attempts to replicate this study. To the best of our knowledge, Alter and Hershfield (2014) is the only published paper that includes an experimental manipulation that influences self-reported search for meaning. Accordingly, we read it with great interest as researchers who study meaning in life and sought to integrate it into our own program of research. In the process of doing so, we conducted a preregistered replication of this work. Our primary aim was to replicate and extend this study. Unfortunately, the results for the extension portion of the study were inconclusive. Nonetheless, the replication portion was successful. Given the high profile of the original paper, recent skepticism of some PNAS papers published by social psychologists (Singal, 2016), and, most importantly, the implications of a reliable way to manipulate people's focus on a fundamental existential concern, we believed independently testing this manipulation was a worthwhile endeavor.

As a secondary aim, we also conducted several exploratory analyses with the data. These exploratory analyses concerned the nature of the dependent variable and potential moderators. With regard to the measure, we were not convinced the items used in Alter and Hershfield's original study tapped into the search for

This article was published Online First June 28, 2018.

Jinhyung Kim and Rebecca J. Schlegel, Department of Psychology, Texas A&M University; Elizabeth Seto, Psychology Department, Colby College; Joshua A. Hicks, Department of Psychology, Texas A&M University.

Study materials, data, and additional results are available on the Open Science Framework webpage at https://osf.io/sy8xz/.

Correspondence concerning this article should be addressed to Jinhyung Kim or Joshua A. Hicks, Department of Psychology, Texas A&M University, 4235 TAMU, College Station, TX 77843. E-mail: jhkim82@tamu.edu or joshua.hicks@gmail.com

meaning, per se. The authors report that the measure is designed to capture the extent to which participants are "preoccupied with seeking meaning in life" and that "most of the items were adapted from the so-called Meaning in Life Questionnaire (MLQ, Steger, Frazier, Oishi, & Kaler, 2006), whereas others were prepared for the present study" (p. 17,067). As frequent users of the MLQ, our perception was that the items used in Alter and Hershfield's study tapped into constructs only tangentially related to the search for existential meaning. We thus explore the dimensionality of the measure in replication Study 1 and the generalizability of the results to more established measures in replication Studies 2 and 3.

Replication Study 1

Method

All studies were conducted in accord with APA ethical guidelines and approved by the Institutional Review Board at Texas A&M University (protocol number: IRB2015-0609D; study title: Personality, Life Goals, and Meaning in Life).

Participants. We recruited a sample of 501 adults through Amazon's Mechanical Turk (MTurk; 258 male, 243 female). All participants were between the ages of 30 and 49 ($M = 36.48$, $SD = 5.35$) and received $0.50 in exchange for their participation. This age restriction was part of our preregistered study plan (available on the Open Science Framework (OSF) web page at https://osf.io/sy8xz/), based on indications in the original data that effects were strongest in this age range (details available at the provided OSF web page), and would thus improve our chance of replicating an effect if one exists. Participants were predominantly White (83%) and non-Hispanic (94%). We excluded four participants' responses from data analyses because they did not complete the writing task as instructed (e.g., one of them wrote "does anybody even read this?").

Materials and procedure. The procedure was almost identical to that used in the original paper, with one main distinction. The original paper used two different control conditions (i.e., think about tomorrow and think about one's next birthday). We, however, only used one control condition (i.e., think about tomorrow) to allow us to increase the sample size for each condition.

After accepting the HIT on MTurk, participants consented and then reported their sex and age. If their age was under 29 or over 50, the survey auto-directed to the end and told participants they were ineligible to participate. Eligible participants were randomly assigned to either the experimental or control conditions. Participants were asked to either describe their day tomorrow ($n = 252$, control condition) or that their 40th (or 50th) birthday was tomorrow ($n = 245$, experimental condition; full prompts and materials available at the provided OSF link).

After completing the manipulation, participants responded to the original 12-item Search for Meaning Questionnaire used by Alter and Hershfield ($M = 4.49$, $SD = .92$, $\alpha = .84$). Example items include "At this moment, it feels important to me to understand which aspects of my life have gone well and which ones have gone less well" and "I measure the quality of my life by how positive an impact I have on other people." All responses were made on a 1 (*strongly disagree*) to 7 (*strongly agree*) scale.

Participants completed five items regarding preferences for experiential purchases (Howell, Pchelin, & Iyer, 2012) relevant to our research. The manipulation did not influence this measure, and it will not be discussed further.

Finally, participants responded to several other demographic items (e.g., relationship status, race/ethnicity), were debriefed, and received their completion code.

Results

Primary analyses. Following our preregistration, we conducted an independent samples t test comparing the conditions using the same measure as Alter and Hershfield. The results revealed a significant difference between the experimental ($M = 4.59$, $SD = .91$) and control conditions ($M = 4.38$, $SD = .92$), $t(495) = 2.57$, $p = .01$, $d = .23$. These results replicated the original paper ($M_{\text{experimental}} = 4.39$, $SD = 0.91$; $M_{\text{control}} = 3.99$, $SD = 0.97$; $t(200) = 2.50$, $p = .002$, $d = .36$).

Exploratory analyses: Dimensionality of measure. To explore the dimensionality of Alter and Hershfield's search for meaning measure, we conducted a series of exploratory factor analyses using the Comprehensive Exploratory Factor Analysis (CEFA 3.04) program (Browne, Cudeck, Tateneni, & Mels, 2010). In the factor analyses, we used Crawford-Ferguson Varimax oblique rotations with maximum likelihood extraction. The factor analyses revealed that the measure used in the original paper is not unidimensional. Both two- and three-factor models fit the data (RMSEA = .082, 90% CI [.070, .094] for two-factor model; RMSEA = .056, 90% CI [.041, .071] for three-factor model) better than a one-factor model (RMSEA = .122, 90% CI [.111, .132]), and there were two factors with eigenvalues greater than one (1.462, 1.142). As such, we extracted two factors from the data. The first factor seemed to capture something similar to what the authors purported (i.e., a preoccupation with thinking about meaning). As shown in Table 1, five of the 12 original scale items clearly loaded on this factor (e.g., "At this moment, I am thinking more deeply about my life than I usually do." and "At this moment, it feels important to me to understand which aspects of my life have gone well and which ones have gone less well."). We named this factor "life reflection" to more accurately capture the content of these items. Five of the remaining items clearly loaded on a second factor. This second factor seemed to clearly differ from the first factor in that it captured the extent to which a person values meaning (e.g., "It is more important for me to lead a meaningful life than to lead a happy life."). We thus computed two subscales from the original measure: The first we refer to as "life-reflection" ($M = 4.79$, $SD = 1.07$, $\alpha = .79$) and the second we refer to as "perceived value of meaning" ($M = 4.37$, $SD = 1.13$, $\alpha = .76$). Two items were eliminated because of their low loadings ($<.40$; see Table 1); however, including them to each factor based on their higher loadings did not change the results (available at the provided OSF link).

An independent samples t test revealed a significant difference between conditions on the life-reflection subscale, $t(495) = 3.03$, $p = .003$, $d = .27$, with participants in the experimental condition ($M = 4.93$, $SD = 1.02$) reporting higher levels of life-reflection than those in the control condition ($M = 4.64$, $SD = 1.10$). By contrast, there was no significant difference between conditions on the perceived value of meaning subscale ($M_{\text{experimental}} = 4.44$, $SD = 1.12$; $M_{\text{control}} = 4.30$, $SD = 1.13$), $t(495) = 1.41$, $p = .16$, $d = .12$ (see Figure 1).

Table 1
Factor Loadings for Alter and Hershfield's Search for Meaning Questionnaire (Replication Study 1)

Item	F1	F2
Factor 1: Life reflection (α = .79)		
1. At this moment, I am thinking more deeply about my life than I usually do.	**.64**	.05
2. At this moment, it feels important to me to understand which aspects of my life have gone well and which ones have gone less well.	**.73**	.05
3. In thinking about my life at this moment, I am taking a very broad view rather than focusing on a few specific moments.	**.46**	.03
4. At this moment, it is important to me to think about how my life has gone so far.	**.76**	.07
5. I feel as though this is a turning point in my life when I can choose to fix things that have not gone well and continue to improve on things that have gone well.	**.47**	.23
Factor 2: Meaning valuation (α = .76)		
6. It is more important for me to lead a meaningful life than to lead a happy life.	.08	**.63**
7. If I could choose to live either a hard and meaningful life or a happy life without meaning, I would choose the hard and meaningful life.	.02	**.75**
8. Life is hard enough as it is without having to worry about making it meaningful.	.18	**−.46**
9. It is important to me to make a meaningful difference in this world.	.14	**.70**
10. I measure the quality of my life by how positive an impact I have on other people.	.14	**.56**
Unfactored items		
11. If I were ever going to write a memoir about my life, this would be a good time to do it.	.32	.18
12. One important measure of a life well lived is that people remember you when you're gone.	.25	.36

Note. Factor loadings over .40 appear in bold.

Re-Analysis of Original Study

To examine the robustness of the differential patterns across our identified subscales, we reanalyzed the original data from Alter and Hershfield's Study 2. Consistent with the results from our replication, the results revealed that the manipulation influenced the life-reflection subscale (α = .83), $F(2, 334) = 12.82$, $p < .001$, $\eta_p^2 = .071$, but not the perceived value of meaning subscale (α = .62), $F(2, 334) = .43$, $p = .67$, $\eta_p^2 = .002$. Post hoc contrasts (i.e., Tukey's test) revealed that the experimental group ($M = 4.79$, $SD = 1.20$) differed from both the birthday ($M = 4.34$, $SD = 1.09$; $M_{difference} = .45$, $p = .02$, 95% CI [.057, .852]) and baseline control groups ($M = 4.01$, $SD = 1.26$; $M_{difference} = .78$, $p < .001$, 95% CI [.420, 1.15]) on the life-reflection scale. With respect to the perceived value of meaning subscale, the experimental group ($M = 4.12$, $SD = .99$) did not differ from the baseline ($M = 4.15$, $SD = .99$; $M_{difference} = .03$, $p = .98$, 95% CI [−.321, .270]) or birthday control groups ($M = 4.04$, $SD = .91$; $M_{difference} = .09$, $p = .79$, 95% CI [−.233, .410]). These results are illustrated in Figure 1.

Replication Study 2

Replication Study 2 was a preregistered replication (available at https://aspredicted.org/w8ns3.pdf) of the Alter and Hershfield's Study 2 that more strongly adhered to their method (i.e., we recruited age groups of 20s and 60s and included both of the original control conditions). We also sought to replicate our own differential patterns for the subscales identified in replication Study 1 and Alter and Hershfield's original study.

Method

Power analysis. An a priori power analysis using the effect size from the original paper ($\eta_p^2 = .035$) revealed that a sample size of 429 would yield .95 power to detect an effect. We thus recruited 450 participants to accommodate any potential exclusions.

Participants. We first recruited 454 adults through MTurk based on our power analysis. However, the data did not yield significant differences between conditions on any of the assessed variables (*p*s > .23; more details available at the provided OSF link). Given that the effect size for the primary dependent variable

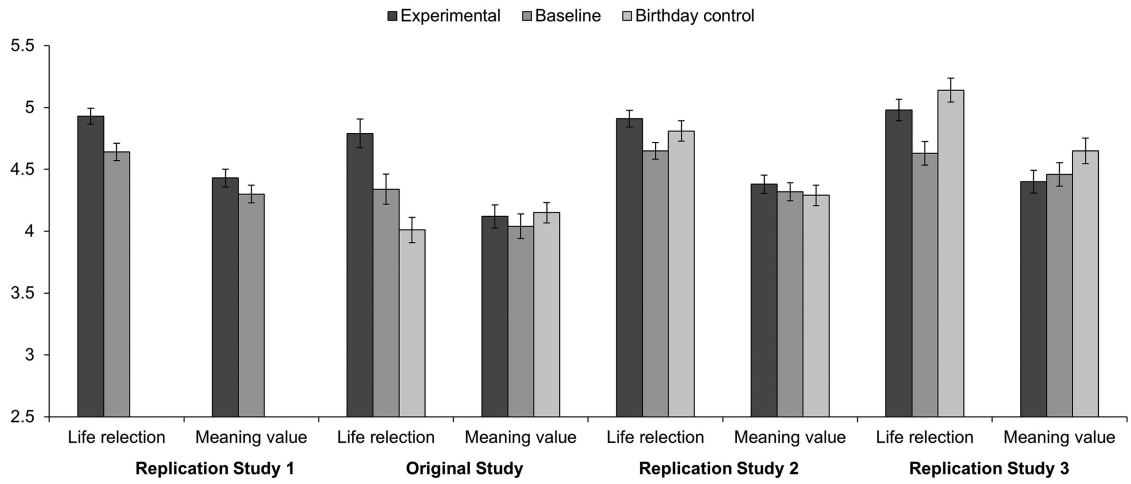

Figure 1. Search for meaning divided by life-reflection and perceived value of meaning subscales between conditions across four studies. Error bars indicate standard errors.

(i.e., Alter and Hershfield's "search for meaning" questionnaire) was found to be much smaller than the estimated one, we collected more subjects, and adjusted our p values based on current recommendations (Lakens, 2014; Sagarin, Ambler, & Lee, 2014). This subsequent data collection also followed our preregistration plan (see https://aspredicted.org/w8ns3.pdf). The subsequent power analysis revealed that adding a sample size of 250 would be appropriate for detecting an effect. Thus, we recruited an additional 248 adults through MTurk, and the final sample size was 702 participants (female = 317, male = 385). Their age ranged from 19 to 70 (M = 34.67, SD = 10.76), which is comparable with Alter and Hershfield's original sample (ranging from 25 to 64; M = 35.70, SD = 10.03). Our analyses below, however, only included participants whose age was between 25 and 64 years old (n = 578).[1] Participants were predominantly White (78%) and non-Hispanic (91%). Three participants' responses were excluded from data analyses because they did not follow the instruction for the writing task (e.g., they typed a few meaningless characters).

Materials and procedure. The study was identical to the Alter and Hershfield's Study 2 except for several measures that were added to the end of the study. As in the original study, participants were randomly assigned to the experimental group (n = 212), the birthday control group (n = 175), or the baseline control group (n = 191).

Upon completing the writing task, participants indicated their agreement with the 12 statements from the Alter and Hershfield's search for meaning questionnaire using a 7-point scale (1 = *strongly disagree*, 7 = *strongly agree*). A composite score of the responses to these items was computed (M = 4.47, SD = .96, α = .85). As in the previous analyses, we also computed life-reflection (M = 4.78, SD = 1.09, α = .78) and perceived value of meaning subscales (M = 4.30, SD = 1.17, α = .77).

We also included the MLQ (Steger et al., 2006) to explore whether the manipulation would have any effect on this more established measure of search for and presence of meaning in life. The MLQ consists of 5 items that assess the search for meaning (e.g., "I am searching for meaning in my life") and 5 items that assess the presence of meaning (e.g., "I understand my life's meaning). Participants completed the MLQ using a 7-point scale (1 = *strongly disagree*, 7 = *strongly agree*), and their responses were averaged to yield their scores of the search for meaning (M = 4.21, SD = 1.52, α = .93) and the presence of meaning (M = 4.67, SD = 1.46, α = .95).

Finally, participants completed the 8-item *reflection* subscale from the Self-Assessed Wisdom Scale (SAWS; Webster, 2003). This reflection subscale captures the extent to which an individual examines and reflects about his or her past experiences in a relation to the present and, by doing so, identifies personal strengths and limitations and ultimately understands the meaning of life. We included this scale to explore whether the manipulation would have an effect on this more established measure of life reflection that has been psychometrically validated (Webster, 2003). Example items include "I often think of my own past," "Remembering my past helps me understand the important things in my life," and "Reviewing my past helps me to have a good perspective of my current concerns." Participants indicated their agreement with each statement using a 6-point scale (1 = *strongly disagree*, 6 = *strongly agree*), and their responses were averaged (M = 4.08, SD = 1.02, α = .91).

Results

Confirmatory analyses. As planned in our preregistration, we performed a one-way ANOVA on Alter and Hershfield's original measure. The results replicated the original finding, $F(2, 575) = 3.41$, $p = .034$, $\eta_p^2 = .012$, although post hoc contrasts revealed that the experimental group (M = 4.60, SD = .95) significantly differed from the baseline control group (M = 4.36, SD = .93; $M_{difference}$ = .24, p = .033, 95% CI [.016, .465]), but not from the birthday control group (M = 4.42, SD = 1.00; $M_{difference}$ = .18, p = .17, 95% CI [−.055, .405]). The baseline and birthday control groups also did not differ ($M_{difference}$ = .07, p = .79, 95% CI [−.171, .301]).

Next, we examined the results for our two identified subscales from replication Study 1. An ANOVA on the life-reflection subscale yielded a significant effect, $F(2, 575) = 4.82$, $p = .008$, $\eta_p^2 = .016$. Post hoc contrasts using Tukey's test found that the experimental group (M = 4.95, SD = 1.07) significantly differed from the baseline control group (M = 4.62, SD = 1.03; $M_{difference}$ = .33, p = .006, 95% CI [.079, .589]), but did not differ from the birthday control group (M = 4.76, SD = 1.17; $M_{difference}$ = .20, p = .18, 95% CI [−.064, .457]). The baseline control group did not also differ from the birthday control group, ($M_{difference}$ = .14, p = .45, 95% CI [−.130, .405]). Consistent with the previous results, an ANOVA on the perceived value of meaning subscale did not find a significant effect ($M_{experimental}$ = 4.39, SD = 1.18; $M_{baseline}$ = 4.23, SD = 1.15; $M_{birthday}$ = 4.26, SD = 1.17), $F(2, 575) = 1.05$, p = .35, η_p^2 = .004 (see Figure 1).

Exploratory analyses: MLQ and SAWS. We then examined more established measures of search for (and presence of) meaning and life reflection. With respect to the MLQ, there was no significant effect on either the search for meaning ($M_{experimental}$ = 4.18, SD = 1.52; $M_{baseline}$ = 4.21, SD = 1.49; $M_{birthday}$ = 4.25, SD = 1.55), $F(2, 573) = .09$, p = .91, $\eta_p^2 < .001$, or the presence of meaning ($M_{experimental}$ = 4.80, SD = 1.39; $M_{baseline}$ = 4.66, SD = 1.47; $M_{birthday}$ = 4.53, SD = 1.51), $F(2, 573) = 1.63$, p = .20, η_p^2 = .006.

Consistent with the results on the life-reflection subscale, we found a significant effect of the manipulation on the reflection SAWS subscale, $F(2, 573) = 2.98$, p = .052, η_p^2 = .01, such that that the experimental group (M = 4.21, SD = .98) significantly differed from the baseline group (M = 3.96, SD = 1.08; $M_{difference}$ = .25, p = .047, 95% CI [.002, .489]), but not significantly from the birthday control group (M = 4.06, SD = .97; $M_{difference}$ = .14, p = .51, 95% CI [−.106, .391]). The difference between the baseline and birthday control groups was not significant ($M_{difference}$ = .10, p = .59, 95% CI [−.146, .353]).

[1] We didn't include this age restriction in our preregistration reports for replication Studies 2 and 3; however, an astute reviewer pointed out that it is needed to more strongly adhere to the methods used by Alter and Hershfield who "specifically excluding college-aged respondents, who are generally prone to seeking meaning based on the nature of adolescence and their tendency to be joining the workforce or entering the final years of formal education, and potential retirees." Analyzing the data using the full data set did not change our interpretations of our findings (i.e., the effect sizes for all analyses were similar). These findings can be found on our OSF page along with exploratory analyses examining how age relates to the effect size of the manipulation.

Replication Study 3

One potential limitation to the interpretation of the results for the previous replication studies is the possibility that the data quality produced by MTurk samples has diminished since Alter and Hershfield (2014) conducted their original study (see Chandler & Paolacci, 2017 for the recent findings suggesting that MTurk participants may lie about their eligibility for participating in studies). This possibility could have contributed to the smaller effect sizes found in the replication studies compared with Alter and Hershfield's (2014) original study. In replication Study 3, therefore, we aimed to directly replicate our previous findings using current recommendations for identifying low quality data (e.g., Aust, Diedenhofen, Ullrich, & Musch, 2013; Oppenheimer, Meyvis, & Davidenko, 2009; Reips, 2010). Namely, we included four additional items to assess whether participants were paying attention and taking the survey seriously. We further restricted the MTurk Worker IDs to help ensure that participants from our previous replication studies did not participate in this study. This study, along with the exclusion criteria, was preregistered (available at https://aspredicted.org/ep2ke.pdf).

Method

Participants. We recruited 449 adults (female = 303, male = 146) through MTurk based on our power analysis (see https://aspredicted.org/ep2ke.pdf). Participants were paid $0.50 in exchange for their participation. Their age ranged from 18 to 71 (M = 37.99, SD = 11.75). Participants were predominantly White (80%) and non-Hispanic (91%). Three participants did not follow the instruction for the writing task (e.g., they wrote "unsure"). Their responses were thus excluded from analyses. Again, the analyses reported in the main text only included participants whose age was between 25 and 64 years old (n = 387); see footnote 1 for more information).

Materials and procedure. The study was identical to replication Study 2 with a few additions (i.e., attention checks). Participants were randomly assigned to the experimental group (n = 135), the birthday control group (n = 119), or the baseline control group (n = 133).

Upon completing the writing task, participants indicated their agreement with statements from the Alter and Hershfield's search for meaning questionnaire, again using a 7-point scale (1 = *strongly disagree*, 7 = *strongly agree*). A composite score of the responses to these items was computed (M = 4.64, SD = .91, α = .81, for all items; M = 4.91, SD = 1.07, α = .72, for the life-reflection subscale; M = 4.50, SD = 1.10, α = .72, for perceived value of meaning subscale).

We again included the MLQ (M = 4.43, SD = 1.43, α = .92, for search for meaning subscale; M = 4.57, SD = 1.43, α = .90, for presence of meaning subscale) and the life-reflection subscale of the SAWS (M = 4.14, SD = .91, α = .87), to explore how this manipulation influences the subjective search for meaning and established life-reflection scales.

Attention checks. To address concerns about the attentiveness of MTurk participants, we included three types of attention-check measures (4 items in total) following recommendations for identifying low quality data from online participants.

First, there were two attention-check items that were embedded among the other questionnaire items (e.g., Alter and Hershfield's questionnaire), instructing participants to select a certain response option (e.g., "For quality control purposes, please select 'Agree' from the responses to the right."). As stated in the preregistration, participants failing to select the indicated responses for both items were excluded from analyses. No one from this age-restricted sample missed both attention-check items (n = 5 for those missing the first item; n = 18 for those missing the second item).

Second, we used a modified item of the instructional manipulation check (IMC, Oppenheimer et al., 2009) from Goodman, Cryder, and Cheema (2013). In this IMC, participants read the paragraph about decision-making research and were asked to select "other" among four options for a question asking "what was the study about?" and also type "decision making" on the line next to "other." The IMC is the most widely used attention-check measure and reported to be effective in improving data quality (e.g., Goodman et al., 2013; Oppenheimer et al., 2009). Our IMC item was embedded among the questions about demographic variables, and there were 27 participants who failed to pass it.

Finally, we directly asked participants to indicate the seriousness of their responses (Aust et al., 2013). That is, at the end of the study before debriefing, participants were asked to indicate whether they have taken part in the study seriously so their data are valid or they did not take part seriously so their data are invalid. Participants were encouraged to be honest by being ensured that they will be paid regardless of their answer. Although it was entirely a self-report measure of attention check, the seriousness check was proven to be an effective way to improve online data validity (Aust et al., 2013; Musch & Klauer, 2002; Reips, 2002). In this study, no one indicated that they did not participate in the study seriously (one participant from the age-unrestricted sample indicated that her participation was not serious).

We also had other exclusion criteria based on the recommendations for data quality improvement (e.g., checking duplicate IP addresses; see Ancillary Analyses section below for more details). By and large, the analyses revealed that there was no noticeable difference in the results no matter what exclusion criteria we separately or simultaneously applied to the analyses. Here, we present the results without applying any exclusion criteria, but refer to the additional materials for the detailed results with various exclusion criteria taken into account.

Results

Confirmatory analyses. As planned in our preregistration, we again performed a one-way ANOVA on Alter and Hershfield's original measure. This revealed a significant difference between conditions, $F(2, 384) = 4.41$, p = .013, η_p^2 = .022; however, the pattern of means was less consistent with predictions. Participants in the birthday control condition reported the highest levels on this measure (M = 4.83, SD = .91), followed by the experimental condition (M = 4.62, SD = .87), and the baseline condition (M = 4.49, SD = .95). Tukey post hoc tests revealed that the only significant difference among those conditions was between the birthday control and baseline conditions ($M_{\text{difference}}$ = .33, p = .009, 95% CI [.068, .601]). The experimental condition did not differ from neither the baseline ($M_{\text{difference}}$ = .13, p = .47, 95% CI [−.130, .386]) nor the birthday control condition ($M_{\text{difference}}$ = −.21, p = .16, 95% CI [−.472, .059]).

An ANOVA on the life-reflection subscale, however, yielded a significant effect in line with predictions, $F(2, 384) = 7.76$, $p < .001$, $\eta_p^2 = .039$. Specifically, the experimental group ($M = 4.98$, $SD = 1.01$) significantly differed from the baseline group ($M = 4.63$, $SD = 1.10$; $M_{difference} = .35$, $p = .022$, 95% CI [.038, .658]), though it did not differ from the birthday control group ($M = 5.14$, $SD = 1.05$; $M_{difference} = -.16$, $p = .67$, 95% CI [−.481, .158]). As for the full measure, the difference between the birthday and baseline control groups was significant ($M_{difference} = .51$, $p < .001$, 95% CI [.196, .823]). Consistent with the previous results, an ANOVA on the perceived value of meaning subscale did not find a significant effect ($M_{experimental} = 4.40$, $SD = 1.07$; $M_{baseline} = 4.47$, $SD = 1.10$; $M_{birthday} = 4.65$, $SD = 1.13$), $F(2, 384) = 1.70$, $p = .16$, $\eta_p^2 = .009$ (see Figure 1).

Exploratory analyses: MLQ and SAWS. Consistent with replication Study 2, there was no significant effect on either the search for meaning, $F(2, 384) = 2.30$, $p = .10$, $\eta_p^2 = .012$, although similar to the pattern for the life reflection subscale ($M_{experimental} = 4.42$, $SD = 1.37$; $M_{baseline} = 4.26$, $SD = 1.49$; $M_{birthday} = 4.64$, $SD = 1.43$), or the presence of meaning MLQ subscales ($M_{experimental} = 4.41$, $SD = 1.42$; $M_{baseline} = 4.69$, $SD = 1.43$; $M_{birthday} = 4.63$, $SD = 1.44$), $F(2, 384) = 1.30$, $p = .27$, $\eta_p^2 = .007$.

In contrast to replication Study 2, however, the results did not yield a significant effect of the manipulation on the reflection SAWS subscale ($M_{experimental} = 4.12$, $SD = .91$; $M_{baseline} = 4.10$, $SD = .95$; $M_{birthday} = 4.22$, $SD = .88$), $F(2, 384) = .60$, $p = .55$, $\eta_p^2 = .003$.

Ancillary analyses. As previously mentioned, it is possible that some of the differences between our findings and Alter and Hershfield's original findings are merely attributable to changing characteristics of recent MTurk participants compared with MTurk participants a few years ago (e.g., they have completed more studies, they are less attentive to the tasks and questions, or read message boards describing specific studies before signing up to participate). To help mitigate these concerns, we reanalyzed all of our data sets implementing various exclusion criteria. In our first set of analyses, we excluded data if an IP address occurred twice in the same data set to account for the possibility that these participants lied about their age to qualify for the study after an initial HIT acceptance revealed they otherwise did not qualify (i.e., they said they were older/younger to complete the study). In the second and third sets of analyses, we excluded participants who did not write very much (e.g., 3 *SD*s below the mean word count) or spent little time in the task (e.g., 3 *SD*s below the mean time spent). Finally, in the fourth set of analyses, as described earlier, we excluded participants who missed attention-check items (e.g., Goodman et al., 2013) and/or indicated that we should not use their data because they didn't take the task seriously (Aust et al., 2013; see https://aspredicted.org/ep2ke.pdf). Again, replication Study 3 was the only study that included these latter items. Across all four sets of analyses, the substantive findings remained unchanged. Complete results and descriptions of these analyses (including number of people excluded) can be found on our OSF web page.

Discussion

Alter and Hershfield (2014) provided evidence for the provocative claim that people are more prone to search for existential meaning when they approach (or think about approaching) a new decade in chronological age. In an effort to test the robustness of their experimental effects, we conducted three high-powered replication studies (total $N = 1466$). Our results generally support the original conclusions of Alter and Hershfield, with some important caveats.

First, the effect size is likely smaller than that reported in the original article. When examining the full meaning-seeking scale, Alter and Hershfield reported a medium effect size ($d = .41$). However, our replication studies found more modest effect sizes on the full scale (ds = .23, .17, and .16, respectively). In fact, although our third replication study ($n = 387$; compared with $n = 337$ for the original study) produced significant overall effect, the difference was not actually statistically different between the experimental and baseline control conditions ($p = .47$), though the means were in line with predictions.

Although it is easy to chalk up these smaller effect sizes to the decline effect in empirical research (i.e., a regression to the mean; Cronbach, 1975; Schooler, 2011), there are likely other variables that influenced the strength of the manipulations. For example, recent participants on MTurk are likely to have participated in many studies (e.g., Stewart et al., 2015) compared with those who completed studies on this platform when the original study was conducted. This experience may increase certain types of response biases. Indeed, recent research shows that MTurk participants are less naïve about experimental manipulations compared with participants who use other online research platforms (Chandler, Mueller, & Paolacci, 2014; Goodman et al., 2013). Moreover, participants in our studies might have been less invested because of our relatively low compensation rate. Although our compensation ($0.50) was higher than Alter and Hershfield's ($0.30), the market rate for completing studies has increased on MTurk (Chandler & Paolacci, 2017; Goodman et al., 2013), perhaps leading some of our subjects to not take the survey as seriously as they might have a few years ago. Each of these possibilities could possibly lead to low quality data, attenuating the true effect size of the manipulation. Finally, it is also possible that people in general are more likely to think about their lives now compared with when the original study took place. Although we do not have specific hypotheses of what might have led to this change, inspection of the means for the experimental versus baseline conditions suggests that, although the means for the experimental group were similar across studies ($M = 4.79$ for the original study vs. average $Ms = 4.96$ for the replication studies), they were notably lower in the baseline condition for the original study versus the replication studies ($M = 4.01$ for the original study vs. average $Ms = 4.63$ for the replication studies), suggesting the possibility that people were more likely to reflect upon their lives at baseline compared with when the original study was conducted.

Although the true effect size is likely smaller than what was found in the original studies, we should note the manipulation developed by Alter and Hershfield is extremely subtle. Most participants only spent a couple of minutes on the writing task (i.e., 1 min and 27.56 seconds as a medium time spent on the writing task for all studies collapsed over condition). The fact that briefly thinking on an upcoming event was enough to make some individuals report wanting to reflect deeply on their life is still remarkable.

A second caveat has to do with the nature of the dependent variable. Based on our findings, it appears that the effect of the manipulation is primarily driven by a subset of the items in the original "meaning-seeking" measure. Our factor analysis revealed the original measure was not unidimensional and that the manipulation only had a significant effect on one of two underlying factors (i.e., life reflection). The other factor (i.e., perceived value of meaning) was unaffected by the manipulation (the same patterns for both subscales were observed in both replication Study 2 and Alter and Hershfield's original study). Although the term "search for existential meaning" is used repeatedly in the PNAS article, the authors do make reference to "self-reflection," "evaluation of life," and "crisis of meaning." This lack of clarity in constructs is not uncommon in the literature (e.g., Park, 2010). Our findings, however, suggests the current manipulation elicits a specific type of meaning search that makes one's life story salient. Consistent with this idea, the results of replication Studies 2 and 3 did not support the idea that this manipulation influences the subjective search for existential meaning as assessed by the more established "search for meaning" subscale within the MLQ. Thus, it may be more appropriate to interpret all the results of Alter and Hershfield's experimental study in terms of life review as opposed to search for meaning, per se. This is a slightly narrower claim than that made in the original paper, but certainly not an uninteresting one. In fact, many researchers have posited that important milestones, such as milestone ages, can lead to a type of "stocktaking" that often includes a deeper reflection about one's life (e.g., Phillips & Smith, 1991). Miron-Shatz, Bhargave, and Doniger (2015) recently found some support for this idea by showing the correlation between physical health outcomes (e.g., BMI) and life satisfaction is stronger for people who recently turned a new decade in life compared with other people. The authors interpret this finding as showing "milestone agers" maintain a more evaluative (vs. emotional) perspective when assessing their lives. Our findings converge with this perspective, to directly support the idea that thinking about milestone birthdays engenders this type of personal stocktaking.

Finally, replication Studies 2 and 3 also revealed that the birthday control condition (i.e., "thinking about your next birthday") did not differ from the experimental condition (i.e., "thinking about you next birthday that would signal entering a new decade"). In fact, there was not a significant difference between the birthday control and experimental conditions on any of our dependent variables. As noted in the preceding paragraph, and at odds with Alter and Hershfield's findings, perhaps thinking about *any* birthday may be enough to elicit an uptick in existential concerns relative to baseline for some individuals.

Of course our replication studies focus on the lone experimental study reported in the Alter and Hershfield's paper. The original paper reports five other studies that examined people *actually* entering a new epoch (i.e., people whose actual age ends in a 9; termed in the paper "9-enders") as opposed to people imagining entering a new epoch. Outside of the lab, the entire year one spends as a 9-ender may feel extra significant and, ultimately, be more impactful on behavior (e.g., leading one to seek an extramarital affair) than being prompted to imagine entering a new epoch in the laboratory setting. Indeed, the effects observed in the other studies reported by Alter and Hershfield may have little to do with birthdays per se.

Although our findings raise many questions, we believe our results converge with Alter and Hershfield's findings to suggest that thinking about the next decade of one's life, and perhaps even any impending birthday, influences the desire to think more deeply about one's life. Although there are likely boundary conditions to this effect (e.g., differential effects of age), these findings warrant future research to both test whether this manipulation influences more than self-reported items assessing this type of existential mindset and illuminate the downstream effects of reflecting on one's life story.

References

Alter, A. L., & Hershfield, H. E. (2014). People search for meaning when they approach a new decade in chronological age. *Proceedings of the National Academy of Sciences of the United States of America, 111,* 17066–17070. http://dx.doi.org/10.1073/pnas.1415086111

Aust, F., Diedenhofen, B., Ullrich, S., & Musch, J. (2013). Seriousness checks are useful to improve data validity in online research. *Behavior Research Methods, 45,* 527–535. http://dx.doi.org/10.3758/s13428-012-0265-2

Browne, M. W., Cudeck, R., Tateneni, K., & Mels, G. (2010). *Comprehensive exploratory factor analysis*. Retrieved from http://faculty.psy.ohio-state.edu/browne/software.php

Chandler, J., Mueller, P., & Paolacci, G. (2014). Nonnaïveté among Amazon Mechanical Turk workers: Consequences and solutions for behavioral researchers. *Behavior Research Methods, 46,* 112–130. http://dx.doi.org/10.3758/s13428-013-0365-7

Chandler, J. J., & Paolacci, G. (2017). Lie for a dime: When most prescreening responses are honest but most study participants are impostors. *Social Psychological and Personality Science, 8,* 500–508. http://dx.doi.org/10.1177/1948550617698203

Cronbach, L. J. (1975). Beyond the two disciplines of scientific psychology. *American Psychologist, 30,* 116–127. http://dx.doi.org/10.1037/h0076829

Goodman, J. K., Cryder, C. E., & Cheema, A. (2013). Data collection in a flat world: The strengths and weaknesses of mechanical Turk samples. *Journal of Behavioral Decision Making, 26,* 213–224. http://dx.doi.org/10.1002/bdm.1753

Howell, R. T., Pchelin, P., & Iyer, R. (2012). The preference for experiences over possessions: Measurement and construct validation of the experiential buying tendency scale. *The Journal of Positive Psychology, 7,* 57–71. http://dx.doi.org/10.1080/17439760.2011.626791

Lakens, D. (2014). Performing high-powered studies efficiently with sequential analyses. *European Journal of Social Psychology, 44,* 701–710. http://dx.doi.org/10.1002/ejsp.2023

Miron-Shatz, T., Bhargave, R., & Doniger, G. M. (2015). Milestone age affects the role of health and emotions in life satisfaction: A preliminary inquiry. *PLoS ONE, 10,* e0133254. http://dx.doi.org/10.1371/journal.pone.0133254

Musch, J., & Klauer, K. C. (2002). Psychological experimenting on the World Wide Web: Investigating content effects in syllogistic reasoning. In B. Batinic, U. Reips, M. Bosnjak, B. Batinic, U. Reips, & M. Bosnjak (Eds.), *Online social sciences* (pp. 181–212). Ashland, OH: Hogrefe & Huber Publishers.

Oppenheimer, D. M., Meyvis, T., & Davidenko, N. (2009). Instructional manipulation checks: Detecting satisficing to increase statistical power. *Journal of Experimental Social Psychology, 45,* 867–872. http://dx.doi.org/10.1016/j.jesp.2009.03.009

Park, C. L. (2010). Making sense of the meaning literature: An integrative review of meaning making and its effects on adjustment to stressful life events. *Psychological Bulletin, 136,* 257–301. http://dx.doi.org/10.1037/a0018301

Phillips, D. P., & Smith, D. G. (1991). Suicide at symbolic ages: Death on

stocktaking occasions. In A. A. Leenaars & A. A. Leenaars (Eds.), *Life span perspectives of suicide: Time-lines in the suicide process* (pp. 81–92). New York, NY: Plenum Press. http://dx.doi.org/10.1007/978-1-4899-0724-0_7

Reips, U. (2002). Internet-based psychological experimenting: Five dos and five don'ts. *Social Science Computer Review, 20,* 241–249.

Reips, U. (2010). Design and formatting in Internet-based research. In S. D. Gosling, J. A. Johnson, S. D. Gosling, & J. A. Johnson (Eds.), *Advanced methods for conducting online behavioral research* (pp. 29–43). Washington, DC: American Psychological Association. http://dx.doi.org/10.1037/12076-003

Sagarin, B. J., Ambler, J. K., & Lee, E. M. (2014). An ethical approach to peeking at data. *Perspectives on Psychological Science, 9,* 293–304. http://dx.doi.org/10.1177/1745691614528214

Schooler, J. (2011). Unpublished results hide the decline effect: Some effects diminish when tests are repeated. *Nature, 470,* 437. http://dx.doi.org/10.1038/470437a

Singal, J. (2016, October 12). *Inside psychology's 'methodological terrorism' debate*. Retrieved from http://nymag.com/scienceofus/2016/10/inside-psychologys-methodological-terrorism-debate.html

Steger, M. F., Frazier, P., Oishi, S., & Kaler, M. (2006). The meaning in life questionnaire: Assessing the presence of and search for meaning in life. *Journal of Counseling Psychology, 53,* 80–93. http://dx.doi.org/10.1037/0022-0167.53.1.80

Stewart, N., Ungemach, C., Harris, A. J., Bartels, D. M., Newell, B. R., Paolacci, G., & Chandler, J. (2015). The average laboratory samples a population of 7,300 Amazon Mechanical Turk workers. *Judgment and Decision Making, 10,* 479–491.

Webster, J. D. (2003). An exploratory analysis of a self-assessed wisdom scale. *Journal of Adult Development, 10,* 13–22. http://dx.doi.org/10.1023/A:1020782619051

Received October 25, 2017
Revision received March 10, 2018
Accepted March 22, 2018 ■