Published in final edited form as:

J Pers Soc Psychol. 2017 December ; 113(6): 939–961. doi:10.1037/pspp0000115.

The Effect of Mood on Judgments of Subjective Well-Being: Nine Tests of the Judgment Model

Stevie C. Y. Yap[1], **Jessica Wortman**[1], **Ivana Anusic**[1], **S. Glenn Baker**[2], **Laura D. Scherer**[2], **M. Brent Donnellan**[3], and **Richard E. Lucas**[1]

[1]Department of Psychology, Michigan State University

[2]Department of Psychology, University of Missouri

[3]Department of Psychology, Texas A&M University

Abstract

Life satisfaction judgments are thought to represent an overall evaluation of the quality of a person's life as a whole. Thus, they should reflect relatively important and stable characteristics of that person's life. Previous highly cited research has suggested that transient factors, such as the mood that a person experiences at the time that well-being judgments are made, can influence these judgments. However, most existing studies used small sample sizes, and few replications have been attempted. Nine direct and conceptual replications of past studies testing the effects of mood on life satisfaction judgments were conducted using sample sizes that were considerably larger than previous studies (Ns = 202, 200, 269, 118, 320, 401, 285, 129, 122). Most of the nine studies resulted in nonsignificant effects on life satisfaction and happiness judgments, and those that were significant were substantially smaller than effects found in previous research.

Keywords

Life satisfaction; subjective well-being; mood; measurement; replication

Subjective well-being (SWB) is an evaluation of an individual's overall happiness and satisfaction with life (Diener, 1984). These evaluations provide an important indicator of quality of life of individuals and the larger society (Diener, Lucas, Schimmack, & Helliwell, 2009). Understanding the nature of SWB and the factors that predict this important construct has both theoretical and applied value. Indeed, governments and public policy makers have begun to recognize the importance of subjective well-being and are increasingly looking to SWB research both to guide policy decisions and provide ways to gauge well-being and mental health at the national level (Diener et al., 2009; Stiglitz, Sen, & Fitoussi, 2009; University of Waterloo, 2011; U. S. Department of Health and Human Services, 2014).

In contemporary research, SWB is most commonly measured using global self-reports. Such measures usually require respondents to make broad, retrospective evaluations of the

Correspondence concerning this article should be addressed to Richard E. Lucas, Department of Psychology, Michigan State University, Psychology Building, 316 Physics Road, East Lansing, MI 48824, lucasri@msu.edu.

Data, along with study and supplemental materials are available on our Open Science Framework page at: osf.io/38bjg.

circumstances of their lives. People are often asked to report on their satisfaction with their lives as a whole, but ratings can also be made about narrower domains of respondents' lives (e.g., satisfaction with health) or judgments of one's own mood or affect. These types of measures offer a relatively efficient and cost-effective method to assess quality of life, and much of what is known about SWB is based on research using global measures.

Some researchers have questioned the validity of global self-report measures by raising concerns about the process by which people make subjective well-being judgments. One of the primary arguments is that global measures are cognitively taxing for respondents. For instance, some scholars have argued that it is difficult for people to evaluate and aggregate across all the relevant aspects and domains of their lives in a short amount of time (Schwarz & Strack, 1999). Thus, it has been proposed that respondents may rely on heuristics to make such global judgments. The use of heuristics opens up the possibility of systematic biases in individuals' responses.

In an influential review, Schwarz and Strack (1999) argued that people use thoughts and feelings that are accessible and relevant at the time of judgment to provide ratings for questions about global well-being rather than conducting a thorough consideration of their life as a whole by searching their memories. This means that temporarily salient factors that should theoretically have no bearing on the actual quality of a person's life may influence and systematically bias their reports. For example, if someone was asked to evaluate the quality of her life immediately after eating an especially delicious meal, she may over-emphasize this pleasant experience relative to other more relevant but less salient features of her life, which in turn would lead to an overestimate of the subjective quality of her life. This process would raise concerns about the reliability and validity of SWB measures. On the other hand, if respondents are actually evaluating their overall quality of life when responding to global measures of SWB, then these reports should be relatively immune to small manipulations of the context in which these measures are administered.

Concerns about the influence of transient factors on judgments of SWB is supported by several studies. For example, Strack, Martin, and Schwarz (1988) asked respondents about their global happiness either before or after asking about their happiness with dating. The correlation between dating happiness and general happiness was much larger when the dating question preceded the general question than it was when the two questions were presented in the opposite order. The interpretation of this result is that participants are more likely to incorporate feelings about dating into their overall judgment when that information has been made salient than when it has not.

Concerns about context effects are especially worrisome when the role of transient moods is considered. To understand why, it is first necessary to consider the theoretical associations between mood and broader judgments of subjective well-being, as the two constructs are clearly linked. Indeed, people's long-term levels of moods (averaged over the course of many days, weeks, or months) are considered to be an important component of subjective well-being (Diener et al, 1999). Yet, moods also fluctuate over time, even varying substantially within a single day or from day to day within a single week (Watson, 2000), and this short-term variability can occur even in the absence of major changes in the

conditions of a person's life. Thus, if these short-term fluctuations in mood affect broader judgments of global well-being, then this might be problematic, as the changes that occurred would not be linked to broader evaluations of the quality of one's life. In turn, this would reduce the reliability and validity of the measures themselves.

In most cases, the effects of these idiosyncratic moods would average out, leading only to attenuated effects that result from the lowered reliability of the measures. However, if moods can also be affected by features of the assessment setting, then these systematic differences could lead to substantial biases in surveys designed to assess the well-being of a population. For example, in one experimental study designed to test the role of mood in life satisfaction judgments, Schwarz and Clore (1983; Experiment 2) asked participants to report their global SWB either on sunny days or rainy days. Respondents in this study reported worse moods when the weather was rainy than when it was sunny. But more importantly, reports of life satisfaction were also affected by this contextual feature, a feature that should be irrelevant for the actual quality of a person's life. Thus, it appeared that respondents' current mood had large effects on the reports that participants provided. This study not only suggested that respondents use their mood at the time of judgment as a heuristic for judging their general well-being, but also that these effects could be large and systematic.

Thinking Critically about Mood, Context, and SWB

The implications of these studies for our understanding and interpretation of well-being research are important, but there are several reasons why these results concerning the impact of mood of reports of SWB should be interpreted with caution. First, although a number of studies have been conducted to examine the association between mood and life satisfaction, many of these studies use extremely small samples. For instance, Schwarz and Clore's (1983) weather study is widely cited and continues to have huge influence in this area of research. Yet, this study used about 14 participants in each experimental condition. Importantly, five of the six conditions within the study were not significantly different from one another, and thus, the effect that they report was driven by a single group of approximately 14 participants. Table 1 provides a list of the studies we could identify that examined the links between manipulated mood and relevant outcomes. Table 1 reports the manipulation that was used, the size of the sample that was included, and the d-metric effect size, when it could be calculated from the reported results (for the procedures we used to calculate these effect sizes, see the supplemental material; this and all other supplemental material can be found on our Open Science Framework Page: osf.io/38bjg). The average per-cell sample size across all studies was 12 participants.

Because the sample sizes are so small, the observed effect sizes in this literature are all necessarily quite large, with Cohen's *d* effect sizes often greater than 1. By our calculation (again, see supplemental material for details), the meta-analytic effect size across the four studies where such effect sizes could be calculated was a *d* of 1.13 for life satisfaction and *d* of 1.50 for a measure of global happiness. In their review of literature, Schwarz and Strack (1999) emphasized that the methodological implications of this work were especially important precisely because the effect sizes were so large. They noted reports of SWB seem "too context-dependent to provide reliable information about a population's well-being" (p.

80) and used language reflecting the large observed effect sizes. For example, they suggested that measures "are *extremely* sensitive to contextual influences" (p. 62, emphasis added) and noted that minor events like finding a small amount of money or having a favorite sports team win a game may "*profoundly* affect reported satisfaction with one's life as whole." (p. 62, emphasis added). However, it is widely known that when underpowered studies are conducted and statistical significance is used as a filter for publication, then published effect sizes will overestimate the true effect (Ioannidis, 2005). Given the almost complete reliance on small-sample studies in this area of research, it is important to replicate these studies using large samples of participants to obtain a more precise understanding of the magnitude of the underlying effects.

Third, although reviews like that presented in Schwarz and Strack (1999) make it seem as though there is robust evidence for the effect of mood on life satisfaction judgments, a close examination of the studies that address this issue shows that the results do not consistently support the idea that mood has strong effects on satisfaction judgments. For instance, although a study by Schwarz (1987), where mood was induced by having participants in the positive mood condition find a small amount of money (a dime), is frequently cited in this context, the difference between the two groups in this study was not statistically significant. Similarly, in a study by Schwarz, Strack, Kommer, and Wagner (1987), participants were called before and after two important soccer games, one a win and the other a tie. Although the authors concluded that the outcome of the game influenced mood, which in turn affected life satisfaction judgments, they only tested the interaction rather than the contrast between participants' scores across the two post-game occasions. Notably, the significant interaction was driven as much by unpredicted and unexplainable pre-game differences as by post game satisfaction, meaning that the post-game scores are likely not significantly different from one another (see supplemental document for further discussion).

Finally, recent work that has taken a large-sample approach to replicating some specific context effects predicted by the judgment model shows that effects sizes are often considerably smaller than those reported in the initial publications. For instance, Schimmack and Oishi (2005) attempted to replicate the basic finding that making specific domains salient before asking questions about global life satisfaction leads to higher correlations between the two measures (e.g., Strack et al., 1988). These authors conducted a meta-analysis of the existing research literature in this area, revealing that the effects of presentation order of domain and global life satisfaction judgments on the global judgments themselves are small when aggregated across studies. These authors concluded that the extant research literature and the analyses of their own data suggest that information that is temporarily salient such as experimental context, presentation order of survey questions and mood have little effect on global SWB judgments (though see Deaton, 2011, for a somewhat larger question-order effect in a much larger sample). Finally, Lucas and Lawless (2013) examined the link between weather and life satisfaction judgments (based on predictions from Schwarz and Clore, 1983) and in contrast to the original study, they found no association between weather and life satisfaction in a sample of over 1 million residents of the U.S. (also see Simonsohn, 2015a). Again, these more recent studies suggest that effect sizes are relatively small (at best) and that direct replications of the original judgment model studies are needed.

The Current Studies

There is debate as to whether subtle changes in context have substantial effects on global reports of SWB. Given the efficiency and widespread use of global measures, resolving this debate is an extremely important goal for research in this field. This paper presents a series of studies that move toward this goal. Each of the studies evaluates the effect of manipulating participants' current mood—one important contextual feature identified in prior research—on subsequent global measures of SWB. Specifically, we present the results of nine studies that both directly and conceptually replicate the procedures of some of the most frequently cited works in the literature on the effect of mood on global judgments of SWB. In some of these studies we replicate, as closely as possible, the methodology of highly cited past work. For instance, in Studies 1 through 4, we attempt to replicate a widely cited finding (Schwarz & Clore, 1983, Study 1) that writing about positive or negative past life events affects global judgments of SWB. Next, in Studies 5 and 6, we conducted two conceptual replications of this study, this time with a different mood induction procedure. In Studies 7 and 8, we conduct close replications of the critical conditions (for our purposes) from the Schwarz and Clore (1983) weather study—those where participants were contacted on days that varied in the pleasantness of the weather and asked about their current life satisfaction. Finally, in Study 9, we conducted a conceptual replication of a widely cited study from Schwarz (1987), where participants who either found a dime on a copy machine or did not find a dime reported differing levels of life satisfaction. It is important to acknowledge from the outset that historical changes have made close replications of this last study near impossible, and thus, the methodology of our conceptual replication is quite different. But the essential element is constant—participants received a small amount of money that should have no bearing on their overall level of SWB. Taken together, these nine high-powered studies represent a broad range of close and conceptual replications of a critical effect in the literature on subjective well-being.

Before describing these studies in detail, it is important to consider two issues that emerge when conducting any replication study. The first of these issues concerns what, exactly, from the original study it is that we are attempting to replicate. For instance, Schwarz and Clore's (1983) second study purports to show that (a) mood affects life satisfaction judgments, (b) the mood that results from relatively mundane events (like an especially nice or especially unpleasant spring day) is enough to substantially shift life satisfaction judgments, and (c) people primarily use their mood as a basis for global well-being judgments when they are not aware that their mood has another cause that would be irrelevant for their life as a whole. In addition, Schwarz and Strack (1999) later suggested that (d) these effects are large enough to undermine the validity and reliability of global self-reports of life satisfaction. It is important to be clear at the outset that our goal is not to test the mood-as-information theory that motivated the original work (the theory that motivated point (c) from above). Thus, we do not include the misattribution conditions from the prior studies, and our results cannot speak to the validity of this broader theory. Instead, our focus is simply on whether standard experimental manipulations designed to affect mood causes changes in life satisfaction judgments and if so, what the size of these effects are when procedures similar to those from the original studies are used.

The second issue concerns the standards used to evaluate replication studies. Like other systematic replication attempts before ours, we acknowledge that there is no single criterion that can be used to determine whether a replication attempt obtains the same results as the original study. In this paper, we consider (a) whether the study obtained a statistically significant effect in the same direction as the original, (b) whether the original result falls outside of the confidence interval of the replication attempt, and (c) whether the obtained result is too small to have been detected in the original study (Maxwell, Lau, & Howard, 2015; Simonsohn, 2015b). In addition, after presenting the individual studies, we consider the meta-analytic average effect size when aggregating results from the original studies (where there is enough information to calculate effect sizes) and the new studies we conducted.

Studies 1, 2, and 3: Life Event Manipulation

Studies 1, 2, and 3 were designed to be close replications of Study 1 from Schwarz and Clore (1983). In the original study, participants were first exposed to a mood induction procedure where they were asked to write about either a positive or negative life event. Next, participants were presented with a series of questions about their mood and subjective well-being. Importantly, the original studies were also designed to assess whether a misattribution procedure could eliminate the impact of mood on well-being judgments, and thus, there were additional conditions designed to affect the extent to which participants used mood when making SWB judgments. Specifically, participants completed the study in a sound-proof room and were either told that the room tended to make people tense, that it made people elated, or participants were given no expectation. Because our goal was simply to evaluate the effect of mood on life satisfaction when no misattribution procedures were used, we did not include these additional conditions. Although this is a difference from the original procedure, there is no theoretical reason to expect that our mood induction procedure (which mirrors that of the “no expectation” condition) would only work in sound-proof rooms (indeed, such a limitation of generalizability would substantially reduce the practical implications of the findings). Furthermore, other studies by some of the original authors have used the life-event procedure without the sound-proof-room component (Strack, Schwarz, & Gschneidinger, 1985). According to the authors, a critical feature of this type of design is that participants are unaware that the mood induction procedure is related to the later well-being questions. Thus, we followed the design of the original study and told participants that the life-event portion of the study was being used to design a life events inventory for college students.

Schwarz and Clore (1983) found that in the no-expectation condition (which the current studies are designed to replicate), participants reported higher life satisfaction and greater global happiness after writing about a positive event than after writing about a negative event. Although the original authors did not provide information that is typically used to calculate effect sizes (such as standard deviations), the raw means, when combined with information about the main-effect F test, can be used to calculate the pooled standard deviation, which in turn can be used to estimate the size of these effects (see supplemental material for more details about these calculations). According to these calculations, the *d*-

metric effect size for life satisfaction was 1.38 (95% CI [0.34, 2.43]). The effect size for a measure of global happiness was an even larger *d* of 2.28 (95% CI [1.08, 3.49]).

Method

Participants—In Studies 1, 2, and 3, participants were undergraduate students at Michigan State University who participated in exchange for partial course credit. As we describe below, participants in Studies 5 through 9 were also students at Michigan State University, whereas those from Study 4 were from the University of Missouri. Participant characteristics for all nine studies are presented in Table 2.

Power—As noted above, *d*-metric effect sizes in the original studies tended to fall in the range of 1 to 2, with an average of 1.13 for measures of life satisfaction and a 1.50 for measures of global happiness. Given concerns about publication bias (e.g., Ioannides, 2005), we assumed that these effects were overestimates of the true effect size, but were unsure what that the true effect size would be. Thus, we chose to power our studies to detect effects that were much smaller than the effects found in the original study, while balancing this concern with limited resources. Thus, in the initial studies, we set out to recruit approximately 100 participants per condition, and then modified this goal in later studies when additional resources were available (e.g., Studies 3, 5, and 6) or as additional design features were added (e.g., Studies 7 and 8). No data analyses were conducted until all data collection was complete.

As a result of these decisions, our studies were well-powered to detect effects of the size found in the original studies, given that most of our studies have over 100 participants per condition. This sample size results in greater than 99% power to detect a *d* of 1.0, over 94% power to detect a much smaller *d* of .5, and 80% power to detect a *d* of .4. In one study (Study 9), we were not able to collect enough participants before the end of the semester to achieve the goal of 100 participants per condition. In addition, Study 4 was collected as part of a broader project and thus, power to detect the original mood induction effect was not an initial consideration. However, these two studies still have more than adequate power to detect effects of the size originally reported, and approximately 77% power to detect a *d* of .5. Furthermore, all of the studies (including Studies 4 and 9) meet Simonsohn's recommendation that replication attempts have at least 2.5 times the sample size of the original study. Of the four original studies from which effect sizes could be calculated, sample sizes for the two relevant conditions ranged from 16 to 28, with an average sample size of 22 participants (11 per condition). Thus, the minimum sample size required to meet this criterion would be 55 participants; all of our studies exceed this criterion.

Procedure—Participants attended in-person lab sessions to complete all measures and study procedures. Upon arrival to the laboratory, each participant was shown to a private cubicle and completed all study questionnaires on computers. Participants first completed the mood manipulation procedure and then completed surveys measuring our key dependent variables. Of particular interest in these studies were the global life satisfaction and global happiness measures, along with assessments of current mood (which always followed the

global measures, on a different screen, so that the participants' mood state would not be made salient before making these global judgments).

Participants were led to believe they would be participating in development of a new measure called the “MSU Life-Event Inventory”, the purpose of which would be to assess events in people’s lives. They were told that they would be asked to describe a particular event from their lives, and that their description would provide the basis for the generation of items in the life-event inventory. Participants were randomly assigned via computer to complete either a positive event or negative event writing task[1]. The complete text of the writing manipulation is presented in the Appendix. In the positive event condition, participants were asked to recall a previous event that made them feel really good. In the negative event condition, participants recalled an event that made them feel really bad. In both conditions, participants were asked to recall the event as vividly as possible, thinking about how the experience made them feel and why. Participants were asked to write about this experience for approximately ten minutes. Following this writing task, participants completed measures of life satisfaction, happiness, and current mood, purportedly to help the researchers select the appropriate response scales for the new life-event inventory. All participants were debriefed regarding the true nature of the study following the completion of the survey.

As we note below, some minor differences in results emerged for the life satisfaction measures (which was the item first presented after the mood induction) as compared to the happiness item (which was administered second). Therefore, in Studies 2 and 3 we tested whether the order of global items might explain the difference in mood effects on happiness and life satisfaction. Accordingly, in Studies 2 and 3 only, participants were also randomly assigned to an item-order condition using the questions about life satisfaction and life happiness mentioned above. In this manipulation, half of the participants saw a question about life satisfaction (“How satisfied are you with your life as a whole?”) followed by a question about their global happiness (“How happy are you with your life as a whole these days?”). The other half of the participants saw the same questions in the reversed order, responding to the question about life happiness prior to the question about life satisfaction. In all studies, the mood measures (which were included as a manipulation check) were assessed after the primary outcomes.

Measures—Global well-being was measured using two items in each study. Specifically, participants were asked how satisfied they were with their lives as a whole (0 = *completely dissatisfied*, 10 = *completely satisfied*) and how happy they were with their lives as a whole these days (1 = *completely unhappy*, 7 = *completely happy*). In addition, two questions were asked as manipulation checks that mirrored those in the original Schwarz and Clore (1983) study: how happy and unhappy they felt in that moment (1 = *not at all happy/unhappy*, 7 = *completely happy/unhappy*), and how good and bad they felt in that moment (1 = *not at all*

[1]We assessed a number of individual differences that were collected prior to the start of the manipulation that we do not analyze or report here. They were collected for the purpose of answering other research questions. All materials are presented on the associated OSF page.

good/bad, 7 = *completely good/bad*). These four questions were included in the original Schwarz and Clore (1983) study.

In addition, in these first three studies, participants responded to at least six questions regarding their mood. Specifically, respondents were asked to report the extent that they felt the following emotions on a 1 (*very slightly or not at all*) to 5 (*extremely*) scale: happy, pleasant, joyful, sad, upset, worried. Their responses were then aggregated into positive and negative mood scores, with the first three items relating to positive and the last three to negative mood. This score showed acceptable reliability, with an average of $\alpha = .91$ for the positive mood score (ranging from .89 to .93), and $\alpha = .73$ for the negative mood score (ranging from .61 to .78).

Results

A summary of the focal results for Studies 1 through 3 is presented in Table 3. Note that the sample sizes vary slightly across the measures in each study assessed due to small amounts of missing data. In addition, Figures 1 and 2 display forest plots for life satisfaction and happiness effects from the original studies by Schwarz and colleagues, the results from all nine of our studies, and the meta-analytic effect size that we describe in more detail below.

Study 1: Life-Events Study 1—As a manipulation check, we tested whether mood varied by condition, as predicted. Participants in the positive event condition reported higher positive mood ($d = .69$) and lower negative mood ($d = -.72$) than participants in the negative event condition following writing manipulation. In addition, participants in the positive event condition also reported that they were significantly happier ($d = .52$) and less unhappy ($d = -.60$) at the time than participants in the negative event condition. Finally, participants also reported that they felt more good ($d = .53$) and less bad ($d = -.54$) in the positive vs. the negative event condition. Thus, these manipulation checks suggest that the mood induction was successful.

For our primary analysis, we examined the effect of the mood induction procedure on the global well-being measures (global life satisfaction and overall happiness). We found no significant effects of the writing manipulation for overall happiness ($d = .10$). However, we did find a significant effect on life satisfaction, with participants in the positive event condition reporting moderately higher levels of life satisfaction than participants in the negative event condition ($d = .32$). Although statistically significant, this effect was smaller than the estimate from the original study ($d = 1.38$ for life satisfaction).

Overall, the results from Study 1 suggest that writing about life events produces effects on mood (with effects sizes in the moderate to large range). Furthermore, because these manipulation checks were included at the end of the study (a design feature that prevents participants' attention from being drawn to the fact that mood was manipulated), we know that the effect on mood lasted long enough to potentially influence the primary outcome measures. However, the effect of mood on global well-being measures was substantially smaller than in prior research, with a small but significant effect for life satisfaction and an effect that was close to zero and nonsignificant for global happiness. Because the life

satisfaction item was always presented first in this study, we examine the role of that presentation order of the global items may play in Study 2.

Study 2: Life Events Study 2—Although the procedure for Study 2 was identical to Study 1, the results of the manipulation check differed to some extent. Results of Study 2 showed that participants in the positive event condition reported a significantly higher positive mood than participants in the negative event condition (d = .34). However, the effect on negative mood was not statistically significant (d = −.21). We also did not find that participants in the positive event condition were higher in their current happiness or unhappiness (d = .23 and d = −.19, respectively) or feelings of good versus bad (d = .13 and d = −.24). Thus, the effects of the writing manipulation on mood were smaller than what we found in Study 1. Although this may seem surprising, it is not entirely unexpected, given that even with moderately large sample sizes, sampling variability can lead to substantial variability in effect size estimates (Cumming, 2011).

Even though there was a significant effect (though reduced in size when compared to Study 1) of the manipulation on positive mood, participants in the positive event condition did not report higher global happiness (d = .11) or higher life satisfaction (d = −.03). This suggests that, again, although writing about a positive event has meaningful effects on mood, it has weaker, and in this study, non-significant effects on global happiness and life satisfaction as compared to the original study.

To test whether item order moderated the impact of the experimental manipulation on subsequent reports of life satisfaction, we varied whether participants first responded to the question about life satisfaction or global life happiness. Results indicated that item order did not affect participant responses for these items. Individuals who were first asked about how happy they were with their lives reported the same level of life satisfaction (M = 6.89, SD = 2.0) as participants who were first asked about how satisfied they were their lives (M = 7.13, SD = 1.62; t(197) = .93, p = .35, d = .13, 95% CI = [−.15, .41]). Similarly, there were no differences in Life Happiness across conditions (t(197) = .50, p = .62). A factorial ANOVA showed that the effect of condition on life satisfaction remained non-significant even when including item order as a predictor (F(1, 195) = .04, p = .84), as did the effect of item order (F(1, 195) = .85, p = .36). Finally, there was no significant interaction between condition and item order (F(1, 195) =.71, p = .40). Similar results were obtained for life happiness (F(1,195) = 0.60, p = .44; F(1,195) = 0.27, p = .61; F(1,195) = 2.64, p = .11; for Mood Condition, Order, and the interaction, respectively).

Taken together with the results of Study 1, Study 2 showed that writing about positive and negative life events can produce effects on individuals' moods, but that the impact of writing about positive or negative events on global happiness or life satisfaction is weaker. The inconsistencies between the results of Studies 1 and 2 on the impact of mood manipulations on life satisfaction led us to conduct Study 3 to again evaluate the robustness of this effect.

Study 3: Writing Study 3—As in Studies 1 and 2, we first conducted a manipulation check, comparing mood reports for those in the positive and negative life-event conditions. Participants in the positive event condition reported more positive mood (d = .59) and less

negative mood (d = −.58) than participants in the negative event condition. This pattern emerged for all of the measures of current happiness or unhappiness (average $|d|$ = .54), with all differences statistically significant (see Table 3).

As in Study 1, individuals in the positive event condition reported higher life satisfaction than individuals in the negative event condition (d = .26). In addition, individuals in the positive event also reported significantly higher levels of global happiness (d = .38). Thus, individuals in the positive event condition judged their global happiness and life satisfaction to be somewhat higher than individuals in the negative event condition. Again, however, the effect sizes in this study were substantially lower than that reported in the original paper, with effects less than one-fifth the size of the original effects.

Consistent with Study 2, there was no indication that responses varied based on the order in which participants were asked to respond to questions. Participants who were first asked about how happy they were with their life reported the same level of satisfaction (M = 6.98, SD= 2.1) as participants who were first asked about how satisfied they were with their life (M = 7.09, SD = 1.9; $t(267)$ = .42, p = .67, d = .06, 95% CI = [−.18, .30]). Differences in life happiness were also not significant, $t(267)$ = .42, p = .67. A one-way ANOVA showed that the effect of condition on life satisfaction remained significant when taking into account item order ($F(1, 265)$ = 4.32, p = .04), whereas the effect of item order remained non-significant ($F(1, 265)$ = .13, p = .72), and there was no significant interaction between condition and item order ($F(1, 265)$ = .29, p = .59). Similarly, for happiness, the effect of condition was significant, $F(1, 264)$ = 9.69, p = .002, whereas the effects for item order, $F(1, 264)$ = 0.27, p = .60, and the interaction, $F(1, 264)$ = .06, p = .81, were not

Study 4: Writing Study

Studies 1 through 3 were designed to be fairly close replications of Study 1 in Schwarz and Clore (1983). Although our goal was to repeat the theoretically important aspects of this study, it is possible that contextual features that the original authors did not consider when discussing the limits of generalizability of their results may affect the outcome of this study. Thus, it is useful to attempt replications that take place in a different context, conducted by different researchers. In the course of conducting these studies, the Michigan State University team learned of a similar independent replication attempt at the University of Missouri. This fourth replication attempt was also designed as a direct replication of Schwarz and Clore Study 1, with features that allowed for an extension. These additional conditions (described in more detail in the online supplement) do not affect the extent to which the conditions below replicate those from Schwarz and Clore, and thus, only the two conditions reflecting a direct replication are discussed in detail. More details are provided in the supplemental material.

Method

Participants—Participants were 118 undergraduates from the University of Missouri who participated in exchange for partial course credit. Age ranged from 18–22, 69% of participants were female, 88% of participants were white, and 11% were African-American.

Design—In the two conditions reported here, participants were randomly assigned to write about a bad life event (negative mood induction) or a neutral event (daily routine; neutral mood induction), before responding to the dependent measures. The purpose of these two conditions was to examine the effect of negative mood on judgments of global life satisfaction and well-being. For a description of the full study design, please visit the Open Science Foundation page specific to Study 4: osf.io/yrcg3 **(link will be open when manuscript is unblinded)**.

Procedure—After arriving to the lab, participants were shown to a computer in a small private room to complete the study. In this attempted replication of Schwarz and Clore (1983), the room itself was deliberately designed to make participants feel uncomfortable while taking the survey, particularly by the introduction of what was ostensibly background ventilation noise (see our OSF page: osf.io/yrcg3). In other conditions not reported here, participants had their attention drawn to the room—thereby encouraging them to attribute their feelings to the room—using a "room survey" which was also a direct replication of Schwarz and Clore (1983). In the two conditions that we report here, participants were not encouraged to attribute their feelings to the room. Notably, the attribution manipulation had no influence on participants' judgments (see our OSF page: osf.io/yrcg3).

Participants first completed the mood manipulation. For the mood induction, participants were asked to write, "As vividly and in as much detail as possible, a recent event that made them feel really bad" (instructions from Schwarz & Clore, 1983). Participants were given 20 minutes to complete the writing task, also consistent with Schwarz and Clore (1983), and they were encouraged to write for as close to 20 minutes as they could. Control participants were asked to write about their daily routine, by being "as detailed as they could" and to list things "in the order that they occurred". Participants were told that the purpose of the study was the creation of a 'life-event inventory', also a direct replication of Schwarz and Clore (1983).

Immediately following the writing task, participants responded to questions about global life satisfaction, followed by subjective happiness, and finally well-being. After this, participants completed a mood manipulation check. Finally, participants answered a few other manipulation check items not relevant to the conditions reported here, and demographic questions. Participants were thoroughly debriefed.

Measures—Unlike Studies 1 through 3, global life satisfaction was measured using the 5-item Satisfaction with Life Scale (Diener, Emmons, Larsen, & Griffin, 1985; α=.89). In addition, participants completed a 4-item subjective happiness scale (Lyubomirsky & Lepper, 1999; α=.86). Well-being was measured using the Bradburn scale of Well-being (Bradburn, 1969), which include a series of dichotomous affect items (α= .65 for the positive affect items, and α =.60 for negative affect items). Finally, current mood was assessed by examining two questions: "How happy (unhappy) do you feel right now?" How good (bad) do you feel right now?" These questions were presented on a 1–7 scale.

Results

Writing Induction—Participants who were asked to write about a negative event followed directions, with no participants being found to have written off topic. Furthermore, topics written about were strongly negative, and included accounts dealing with death, isolation, divorce, rape, and violence. Participants wrote for an average of 14.87 minutes (SD= 7.15) in the negative condition and 13.56 minutes (SD= 6.15) in the neutral (control) condition. Table 4 shows differences in responses across conditions. Participants who wrote about a bad event reported feeling somewhat less happy (M=4.74, SD=1.29) than those in the neutral (control) condition (M =4.89, SD=1.06), but this trend was not significant, $t(116)$=0.72, p=.47, d=0.13. Similarly, participants who wrote about a bad event reported feeling somewhat worse (M=4.97, SD=1.34) than those in the neutral condition (M=5.21, SD=1.11), but again this trend was not significant, t(116)= 1.07, p=.29, d=0.20.

Dependent Measures—There was no effect of writing task on responses of life satisfaction, $t(116)$=1.13, p=.26, d=0.21. This was also the case for subjective happiness ($t(115)$=−.35, p=.73, d=−0.07) and the Bradburn positive ($t(116)$=0.63, p=.53, d=−0.12) and negative affect ($t(116)$=−1.21, p=.23, −d=−0.22) items. It is important to note, however, that although the effect sizes were not significant, the effect size for life satisfaction was in the range of the effect sizes found in Studies 1 through 3. Although this study was adequately powered to detect effects that were substantially smaller than those in the original study, it was not adequately powered to detect the small effects found across the previous three studies.

One concern that can be raised about this study is that the writing manipulation did not significantly influence mood, in spite of the fact that participants wrote for almost 15 minutes on average and generated considerably negative content. One reason could be that this study included only negative and neutral mood conditions, and not a positive mood condition. However, in the original Schwarz and Clore (1983) study, writing about a negative event had a “more pronounced effect on subjects’ mood than did the instruction to think about positive events” (p. 516), and participants in the positive writing group did not report higher mood compared to control group, which is consistent with the apparent truism that bad is stronger than good in terms of psychological impact (Baumeister, Bratslavsky, Finkenauer, & Vohs, 2001). These considerations justified the choice of the negative vs. neutral mood conditions as a reasonable approach for the present study.

Studies 5 and 6: Video Studies

Studies 1 through 4 represented fairly close replications of a highly influential study that examined the links between induced mood and global reports of subjective well-being. However, the conclusion from the original study was not simply that writing about positive life events leads to elevated reports of global life satisfaction. Instead, the broader point was that current mood itself influences these judgments. Thus, any mood induction (at least one that was presented in such a way that respondents would be unlikely to consciously link the induction to the later judgments) should also lead to differences in global measures of subjective well-being. Thus, the goal of Studies 5 and 6 was to conceptually replicate these

results using alternative mood induction procedures that have in past studies been shown to be effective.

Method

Participants—As in Studies 1, 2, and 3, participants were undergraduate students at Michigan State University who participated in return for course credit. Sample sizes and participant characteristics are presented in Table 1.

Procedure—Upon arrival to the laboratory, participants answered a number of questionnaires unrelated to the current study. Following this, each participant was told that he or she would be watching one or two video clips, ostensibly pre-testing the clips for an unrelated study. They were informed that these videos would be followed by questions about the content of the videos to encourage participants to pay attention (the text of these instructions can be found in the Appendix). In Study 5, all participants first watched a neutral video selected randomly by the computer from a pool of 10 emotionally neutral videos identified and pretested by our research team. For example, one video in this pool discussed the history of Roman forums, and another was an informational video describing how bananas grow. Following this neutral video, participants were asked to respond to three yes or no questions: whether they have viewed the video before, whether they learned anything, and whether they would recommend the video to a friend. These questions were included to support the cover story that we were pre-testing the videos for later use. Participants were then randomly assigned to watch either a happy video clip or sad video clip, followed by the same questions asked after the neutral video. In Study 6 participants only watched a happy or a sad video (i.e., they did not watch a neutral video).

The happy or sad videos were initially identified by our research team, and subsequently pre-tested to create two pools of 5 each. Participants viewed a video from one of these pools randomly selected by the computer (depending on their experimental condition – which was also randomly assigned via computer). Examples of videos participants viewed in the happy video condition included video compilations of laughing babies, an orangutan and a dog playing together, and an inspiring movie of kids playing soccer. Examples of videos participants viewed in the sad video condition included the first four minutes of the animated film “Up,” which depicts the death of the main character’s spouse and the loneliness associated with his widowhood, and a sad scene from the movie “Click”. Full list of videos is available in the Appendix. The videos were approximately 3–5 minutes long.

Following the videos, participants responded to the same measures of life satisfaction and life happiness administered in Studies 1–4. They also responded to the 5-item Satisfaction with Life Scale (SWLS; Diener et al., 1985), and a measure of current mood. In Study 5, participants completed basic demographic questions that asked about their gender, age, and ethnicity prior to questions about SWB and mood, whereas in Study 6, the demographic survey followed the SWB and mood questionnaire.

Results

Study 5: Video Study 1—As in Studies 1 through 4, we first conducted a manipulation check to ensure that the mood induction procedures successfully affected mood. Consistent with expectations, participants who viewed a happy video reported higher positive mood (d = .30) and less negative mood (d = −.32) than individuals who viewed a sad video, suggesting that this video mood manipulation generated mood effects for participants in the expected direction and that the induced mood lasted through the end of the study. However, our primary analyses (presented in Table 5) show that participants in the happy video condition did not report higher global happiness than participants in the sad video condition (d = .04). There was also no significant difference in life satisfaction between participants in the happy and sad video conditions, either when measured with a single-item (d = .11), or when measured with the SWLS (d = .17).

Thus, in this study, watching happy or sad videos affected mood. That is, participants reported higher positive mood and lower negative mood following the mood manipulation. However, the effects on life satisfaction or global happiness were not statistically detectable. To evaluate the robustness of this video effect, we conducted Study 6, a direct replication of this experimental procedure in Study 5 with a different sample of participants.

Study 6: Video Study 2—As in previous studies, we first conducted a manipulation check. Again, participants in the happy video condition reported higher positive mood (d = .59) and lower negative mood (d = −.50) than participants in the sad video condition. These mood effects are larger than in Study 4, and comparable to the largest effects found in Studies 1, 2, and 3. However, despite the larger effect of the induction on reports of mood, participants across conditions again did not differ in their reports of global happiness (d = .20). They also did not differ in their reports of life satisfaction when measured with either the single-item (d = −.03) or the SWLS (d = .04). Again, the results of this study indicate that although mood manipulations appear to influence participants' reported mood, their effects on life satisfaction or global happiness are much smaller.

Study 7: Weather Study I

The goal of Study 7 was to replicate Study 2 of Schwarz and Clore (1983), which focused on the impact of weather on life satisfaction judgments (under the assumption that weather can be a naturally occurring mood induction). Specifically, participants in the original study were called on either a sunny spring day or a cloudy spring day and asked about their life satisfaction.[2] We attempted a conceptual replication in the Spring of 2013 by comparing the impact of a nice spring day (with nice being defined as a combination of warm and sunny) versus a cold, cloudy, and rainy day on a daily report of subjective well-being.

[2]Few details were included in the original study about the specific days on which the study was conducted, or the specific weather conditions that occurred on those days. This makes it difficult to precisely evaluate the similarity in weather conditions across the studies. However, certain pieces of information could be gleaned from the original report and a later paper that discussed the impact of the original study on the psychological literature (Schwarz & Clore, 2003). See the supplement available at our OSF page for a more detailed comparison.

Method

Procedure—Participants signed up on-line for a two-part study involving an initial on-line session session and an on-line follow-up, which was to occur at some point in the future. The total number of participants in the initial pool was 461. Participants completed a questionnaire (via computer) that included questions about their personality and subjective well-being during the lab session. These initial sessions took place between March 15th and March 20th of 2013. The participants were then randomly assigned by the survey software to a good-weather or a bad-weather "re-contact" condition (208 were assigned to the good condition and 253 to the bad condition). We split the re-contact days to two good (April 4 and 16) and bad days (March 20 and April 10) to spread data collection. Participants were then sent recruitment emails on relatively good or relatively bad weather days and asked to report on their global well-being and current mood in an online survey. If participants selected for their first weather day did not respond, they were contacted again on the second day. For example, if a participant in the good condition was selected for the first good day but failed to respond, she/he was contacted on the last good day (this happened for 42 participants in the bad condition and 56 participants in the good condition).

Of the 461 participants in the initial pool, 285 completed the online follow-up survey (62%). There was non-random attrition with 117 of the 208 participants assigned to the good days participating in the online survey (56%) versus 168 of the 253 participants assigned to the bad days (66%); $X^2 = 4.986$, $df = 1$, $p = .026$. However, the two groups did not differ in terms of baseline scores at $p < .05$ (see Results).

For the target days, we selected days that were forecast to be at least 10 degrees Fahrenheit warmer (or colder) and relatively less (or more) cloudy than the previous days. Table 6 shows the dates and the actual temperature, precipitation, and the extent of cloud cover on target days and preceding days. It is important to note that although the average cloudiness on the second good day is similar to those on the days before, the cloudiness and rainfall occurred in the early morning hours. The morning and especially the afternoon were clear and dry, which made the day an appropriate one for testing.

To get a better sense of the actual mood that people experienced on the target days, participants also completed an on-line Day Reconstruction Method survey (modelled on Anusic, Lucas, & Donnellan, in press). In this survey, participants were asked to divide their day into distinct episodes reflecting different activities. They were then asked to reconstruct that day, describing exactly what they were doing and how they were feeling during each episode. Average affect over the day was calculated from this survey. Sample sizes and participants' demographic characteristics can be found in Table 2.

One difference between our study and the original study was that we recruited participants and asked them to complete a baseline survey before re-contacting them to complete a survey on the target weather day. Although no participants were asked to complete the exact same measures used in the target-day follow-up, participants were asked to complete baseline measures involving subjective well-being and trait affect (along with the facets of the Big Five which were not analyzed here). Our goal in administering this measure was to be able to control for pre-existing differences when examining the effect of weather,

allowing for greater power to detect an effect. We test for potential biasing effects of elements of this procedure in Study 8.

Measures

Satisfaction With Life Scale—The 5-item Satisfaction With Life Scale (SWLS; Diener et al., 1985) was used to assess baseline SWB. The SWLS asks participants to rate their agreement with statements such as “In most ways my life is close to my ideal” and “I am satisfied with my life” on a 7-point scale (1 = Strongly disagree, 7 = Strongly agree).

Trait affect—During the initial survey, participants were asked to rate their affect “in general, or on average”. They reported how frequently they experienced each of the listed emotions or feelings on a scale from 0 (Almost never) to 6 (Almost always). Positive affect items were *happy, satisfied*, and *meaning*. Negative affect items were *frustrated, sad, angry, worried, tired*, and *pain*.

Single-item global well-being items—On the target bad/good weather day participants were asked three global well-being questions similar to those used in Schwarz and Clore (1983): “How happy do you feel about your life as a whole?” (0 = Completely unhappy, 10 = Completely happy), “Thinking of how your life is going, how much would you like to change your life from what it is now?” (0 = Change a very great deal, 10 = Not at all), and “How satisfied are you with your life as a whole?” (0 = Completely dissatisfied, 10 = Completely satisfied). The measures differed from those in the original in the number of response options provided (our measures included 11-point scales; the original studies used 10-point scales).

Mood—Current mood was assessed on the target day by asking participants about the extent to which they felt each of the listed emotions at that moment. Positive mood items were *happy, pleasant*, and *joyful*. Negative mood items were *sad, upset*, and *worried*. Ratings were made on a 5-point scale (1 = Very slightly or not at all, 5 = Extremely).

The DRM task—Following the target day, participants completed a modified Day Reconstruction Method (DRM) (Anusic et al., in press), during which they were asked to reconstruct the previous (target) day. They were guided to first recall the time they woke up and the time they went to sleep on the previous day. Then they were asked to reconstruct their previous day by breaking it up into distinct episodes (e.g., having breakfast, driving to work), reporting the times each episode began and ended, and making any notes about the details of the episodes. Following this, we randomly selected three of the listed episodes, and for each of the episodes we asked participants what they were doing, who they were with, and how they felt during the episode. Finally, participants answered some addition questions about their day overall. The questions that we focus on from the DRM are listed below.[3]

[3]When initially asked about the previous day, three participants incorrectly reported the day of the week that corresponded to the target day. Their responses were excluded from the analyses involving the DRM because it was unclear which day they were reconstructing.

DRM-based affect: For each of the rated episodes, participants were asked about the feelings they experienced during the episode. The emotions listed were identical to those used in the trait affect scales, and participant made their ratings on a 7-point scale (0 = did not experience feeling, 6 = feeling was very important part of the experience). We computed a positive and a negative affect score for each episode, and then averaged these scores across all rated episodes. Thus, there was a single positive and a single negative affect score for each day reconstructed through the DRM.

Day satisfaction: Participants were also asked how satisfied they were with their previous day (i.e., the reconstructed day). Participants made their ratings on a scale ranging from 0 (Completely dissatisfied) to 10 (Completely satisfied).

Results

Random Assignment and Attrition—Participants were assigned to weather condition using the randomizer function in Qualtrics survey software (the option to create equal-sized groups was not selected). We tested whether there were baseline differences between the groups in terms of scores on the SWLS and trait affect. There were no statistically detectable differences (smallest $p = .212$ for Negative Affect). Given the non-random attrition across conditions, we also compared whether there were differences in baseline characteristics for participants who completed the daily survey versus those that did not. There were no statistically detectable differences (smallest $p = .313$ for SWLS). We also tested for an interaction between condition assignment and whether participants completed the daily survey for each of the three baseline variable using a 2 by 2 ANOVA framework. None of these terms were statistically significant (smallest $p = .125$ for SWLS). These results reduced concerns related to randomization and attrition about the internal validity of the study.

Main Analyses—Table 7 shows the means of dependent variables across the experimental groups. The most notable finding is that we found no statistically significant effects of weather on global SWB ratings. Weather effects on global evaluations of life (life satisfaction, happiness, and desire for change) were typically small, never exceeding a d of | 0.10|. Second, the largest effects were observed for emotional recall during the DRM, though even these were not significantly greater than zero. Finally, participants who recalled a good weather day for their DRM did not consistently report being more satisfied with their day.

It may be difficult to detect between-group differences in well-being because there are substantial pre-existing differences in well-being among individuals that may be overshadowing the impact of weather. We explored this idea by statistically controlling for pre-existing individual differences in well-being. We repeated our weather analyses including both SWLS and trait affect variables. Controlling for initial levels of well-being did not change the conclusions about the effects of weather on ratings of well-being despite having potentially more statistical power (see online Supplement). Thus, Study 7 suggests that the effects of weather on global well-being judgments are, at best, very small. The null results for weather are generally consistent with the results of Lucas and Lawless (2013),

though the current study adds by more closely replicating the original design of Schwarz and Clore (1983).

Study 8: Weather Study II

The goal of Study 8 was to replicate Study 7 in a different semester. One issue is that it can be difficult to find the "best" day to test the ideas in Schwarz and Clore (1983) given that the specific weather conditions that would be predicted to lead to relatively large differences in mood are relatively rare in any given spring in the Midwest. Moreover, we were concerned that completing baseline measures of well-being might have biased participant responses in Study 7. The reason for including them was to be able to control for pre-existing differences in well-being when examining the effect of weather, allowing for greater power to detect an effect. However, one concern is that those participants who were asked to think about their SWB in this initial session may simply remember that judgment later, which could actually reduce the impact of weather on the judgment.

According to Schwarz and Strack's (1999) discussion of findings of the judgment model, this should not be a concern. For instance, they noted that even within a single experimental session, two assessments of the same exact measure were not strongly correlated, precisely because participants could be influenced by contextual factors that changed within a single hour-long session. Thus, it would be unlikely that a different set of related questions assessed weeks before the target assessment would influence these final judgments. However, to test this possibility, we extended our design in Study 8 by randomly assigning participants at the initial session to either complete a global life satisfaction measure or a similar-length unrelated measure about their artistic interests. If participants relied on their memory when rating their life satisfaction, then any effects of weather should be stronger in the artistic interest condition (this manipulation did not seem to influence the results).

Method

Procedure—As in Study 7, participants signed up on-line for a two-part on-line study involving an initial session and a later follow-up. The total number of participants in the initial pool was 374. Participants completed a questionnaire (via computer) that included questions about their personality and subjective well-being. These initial sessions took place between February 26th and March 18th of 2014. The participants were then randomly assigned by the survey software to a good-weather or a bad-weather "re-contact" condition (192 were assigned to the good condition and 182 to the bad condition). One additional person was not able to complete the first survey due to technical difficulties but participated in the follow-up session and was randomly assigned to the bad-weather condition. We simplified the design from Study 7 by only picking one good and one bad weather day rather than trying to find two good weather days per each condition (March 31 and April 3, respectively). As in Study 7, participants were sent recruitment emails and asked to report on their global well-being and current mood in an online survey.

Of the 375 participants in the initial pool, 260 completed the online follow-up survey (69%). Unlike Study 7, there was no evidence of non-random attrition, with 136 of the 192 participants assigned to the good days participating in the online survey (71%) versus 124 of

the 183 participants assigned to the bad days (68%). The Pearson Chi-Square test statistic was 0.416 ($df = 1$, $p = .519$). One possible explanation for the difference in attrition across studies is that the good and bad weather days in Study 8 were spaced within the same week and thus at a more constant interval between the initial survey and the online survey than in Study 7. Table 6 shows the dates and the actual temperature, precipitation, and the extent of cloud cover on target days and preceding days. As in Study 7, participants also completed an on-line Day Reconstruction Method survey (modelled on Anusic, Lucas, & Donnellan, in press).[4] Sample sizes and participants' demographic characteristics can be found in Table 2.

Measures

Measures in Study 8 were the same as Study 7 with two exceptions.

Pretest - Satisfaction With Life Scale versus Artistic Interest Manipulation— The 5-item Satisfaction With Life Scale (SWLS; Diener et al., 1985) was included for one pretest condition. Participants in the other pretest condition completed five items from an openness to experience measure (IPIP-300; Goldberg et al., 2006) that assessed artistic interest, rather than the SWLS. This scale asked participants to rate their agreement with items "I like music," "I love flowers," "I enjoy the beauty of nature," "I do not like art," and "I do not enjoy watching dance performances" on a 7-point scale (1 = Strongly disagree, 7 = Strongly agree).

Mood Survey Condition Manipulation—During the course of the study we determined that the week of March 31 would be a good opportunity to test weather effects. One complicating issue, however, stemmed from the performance of the school's basketball team in the Men's National Collegiate Athletic Association's Basketball Tournament (i.e., March Madness). That year Michigan State University lost to the eventual tournament champion (University of Connecticut) in the Elite 8 round on March 30 despite having the lead at halftime (Final Score: 60 to 54; source: http://www.ncaa.com/game/basketball-men/d1/2014/03/30/uconn-michigan-st). We were concerned that that this unpleasant event, which occurred the day before the pleasant weather condition, might negatively impact reports of well-being.

Given the difficulty in finding days with weather that seemed to be a match for those in the original study, we decided not to postpone the follow-up sessions. Instead, we built in additional experimental features that could potentially address this concern. Specifically, we randomly assigned online survey participants to complete either the version administered in Study 7 or to a version in which they also complete additional questions at the start of the survey. Based on the idea that making the cause of participants' mood salient would lead them to discount the informational value of that mood (Schwarz and Clore, 1983), we asked questions about the weather and the game. Specifically, participants indicated whether they watched the game: "Did you watch the MSU men's basketball game on Sunday?" (Yes/No); and 95 out of 130 participants answered yes (73%). Participants indicated their affective reactions to the game – "How did you feel about the outcome of the game?" (1 = Very

[4]Five participants incorrectly reported the day of the week that corresponded to the target day and were excluded from the analyses involving the DRM.

negative to 5 = Very Positive). The mean response was 1.91 (SD = .772 with a modal response of negative given by 47% of the participants). As a check on the weather manipulation, participants also rated the weather – "How would you rate today's weather?" (1 = Very Unpleasant to 5 = Very Pleasant). As expected, there was a substantial mean difference (Bad Weather: $M = 1.97$, $SD = .80$, $n = 61$; Good Weather: $M = 4.62$, $SD = .67$, $n = 69$; $t = 20.714$, $df = 128$, $p < .001$).

Results

Random Assignment and Attrition—Participants were assigned to weather condition using the randomizer function in Qualtrics survey software. We tested whether there were baseline differences between the weather groups in terms of scores on the SWLS and trait affect for participants who completed those items. There were no statistically detectable differences (smallest $p = .404$ for Negative Affect). We also compared whether there were differences in baseline characteristics for participants who completed the daily survey versus those that did not. There were statistically detectable differences for the SWLS ($t = 2.198$, $df = 195$, $p = .029$; those with daily survey data were higher) and Positive Affect ($t = 2.643$, $df = 372$, $p = .009$; those with daily survey data were higher) but not Negative Affect ($t = 0.514$, $df = 372$, $p = .608$). There was no effect of the SWLS versus Artistic Interest manipulation on daily survey responses (minimum $p = .097$ for life satisfaction).

Main Analyses—We first conducted a 2 by 2 ANOVA to evaluate the impact of the initial survey items about the basketball game and weather on the key dependent variables (Life Happy, Life Change, Life Satisfaction, Positive mood, and Negative Mood). There was only one statistically significant effect at $p < .05$. The effect of survey condition was detectable for Life Happy item (F = 7.008, $df = 1, 256$, $p = .009$). The effect of weather and the interaction were not statistically significant ($p = .893$ and .332, respectively). Participants who completed the initial survey items about the basketball game and weather scored lower than those that did not (No Questions: $M = 7.70$, $SD = 1.42$, $n = 130$; Questions: $M = 7.20$; $SD = 1.65$, $n = 130$). This effect ran counter to our expectation that making a potential source of mood salient would lead participants to discount that information. No other effects were detectable (smallest $p = .08$ for the effect of Survey condition on Life Satisfaction).

Given this result, we believed that the fairest test of the weather effect (and the most relevant comparison with past work) would be to restrict the analyses to the group that did not complete the additional survey items (results for the full sample can be run using the analysis script in the supplemental material; notably, there were no significant differences across the good and bad weather conditions for Life Happy or Life Satisfaction when the full sample was used). It is important to note at the outset that this complicates the interpretation of these results, as the negative event (the basketball loss) could potentially wipe out the positive effects of the pleasant weather. However, the original effect in Schwarz and Clore (1983) was driven not by a positive reaction to the pleasant weather, but by participants' negative reaction to the unpleasant weather. Thus, it is still useful to compare weather conditions in this study. In addition, as there is no reason to expect differences in this population from one year to the next, in supplemental analyses we also explicitly compare means across Study 7 and 8.

Table 7 shows the means of dependent variables across the weather groups for those who did not complete the extra three survey items. As in Study 7, we found no statistically significant effects of weather on global SWB ratings. Weather effects on global evaluations of life (life satisfaction, happiness, and desire for change) were typically small and counter to predictions given that (if anything) participants responding on the bad weather day scores higher than those responding on the good weather day. Given the reduced sample size, we do not report results controlling for pre-existing SWLS scores but controls for trait affect did not alter these conclusions (see supplemental material). Thus, as with Study 7, Study 8 suggests that the effects of weather on global well-being judgments are, at best, very small.

As noted above, one concern about the interpretation of Study 8 is that the pleasant weather day followed a loss by the MSU Men's Basketball Team. Thus, the effects of the pleasant weather condition are somewhat ambiguous because this condition combined pleasant weather with the potentially unpleasant experience of the loss. However, the unpleasant weather condition, which occurred just a few days later, is much less ambiguous. It actually combines two unpleasant circumstances: the cold and rainy weather and the recent loss (though the impact of this loss may have faded by this time). Furthermore, in the original Schwarz and Clore (1983) study, it was the negative condition that drove the effect, which means that a replication of the effect should show reduced happiness and life satisfaction on the bad weather day. Because there is no reason to expect differences from year to the next, we plotted the means from the pleasant and unpleasant weather conditions from Study 8 against those from Study 7. As can be seen in Figure 3, the means for happiness and life satisfaction are virtually identical across all conditions in the two studies. This provides further support for the idea that weather did not affect the global well-being measures in these studies.

As a final test, we examined whether those who completed the SWLS in the initial session were differentially affected by the weather manipulation as compared to those who instead completed the unrelated artistic interests scale. Specifically, we regressed each outcome variable on a dichotomous weather variable, a dichotomous SWLS/artistic interest variable, and their interaction. If completing the SWLS reduces the impact of weather on the later measures, this interaction should be significant, and the effect of weather should be larger in the artistic-interests condition than in the SWLS condition. To maximize power, we included all participants from the follow-up session, regardless of whether they received the pre-survey questions about the game and the weather. Detailed results are not presented here but can be obtained by running the analysis script supplemental materials. Most importantly, there was no interaction between completing the SWLS in the initial session and weather condition on either the Life Happy or Life Satisfaction measure.

Study 9: Quarter Study

As a final attempt at examining the robustness of mood effects on global well-being judgments, we focused on another widely cited study that purports to show substantial effects of mood on life satisfaction judgments. Specifically, in an early study, Schwarz (1987) manipulated mood by having participants find a small amount of money. Participants were instructed to use a coin-operated photocopier to make copies of the questionnaires they

were to complete as part of the study. Participants were randomly assigned to conditions where they either found a dime on the copy machine or they did not. This study follows in the tradition of previous work, which showed that receiving small amounts of money can lead to surprising boosts in positive mood (e.g., Isen & Levin, 1972, Isen & Simmonds, 1978). Importantly, although life satisfaction scores were substantially higher in the dime condition than in the non-dime condition (with *d*s around .80; see supplemental material), the difference between the two was not statistically significant using conventional alpha levels, thereby undermining the statistical validity of the study. Despite this lack of a significant effect, this study has often been cited as important evidence for the role that mood plays in what are supposed to be relatively stable judgments about the quality of a person's life (e.g., Kahneman & Krueger, 2006). In any case, because of concerns about the statistical validity of the original study, there are reasons to be cautious in our expectation that receiving a small amount of money will lead to differences in well-being judgments.

In addition, although Study 9 was designed as a conceptual replication of Schwarz's (1987) dime study, it was not possible to directly replicate the original procedures. In the original study, participants were asked to use a coin-operated photocopier. We could not replicate this procedure because coin-operated photocopiers are far less common in academic settings in 2010s than in the 1980s. Additional pilot attempts to develop procedures in which participants would serendipitously discover small amounts of money (presumably left in a room by a previous participant) led to considerable suspicion, or participants would simply pass up the opportunity to take the small amounts of money that were left for them. Thus, we designed a novel experimental procedure aimed providing participants with small, unexpected amounts of money.

Method

Participants—As in most of the earlier studies, participants were undergraduate students at Michigan State University, who participated in return for partial fulfillment of a course requirement. Details about the sample are presented in Table 2.

Procedure and Measures—Participants either received a quarter or did not receive a quarter on the basis of the roll of a die. Participants were told that the goal of the study was to replicate a highly publicized recent study examining extra-sensory perception (ESP) among college students. Research assistants asked each participant to make three guesses about which number would roll on a standard six-sided die. They were also told that, as a small incentive, they would receive 25 cents if any of their three numbers appeared, otherwise they would receive no money (the equivalent value of a dime in the mid-1980s was approximately 25 cents in 2012). Participants were asked to concentrate on their chosen three numbers for a few seconds and then roll the die once. With a fair die (and absence of ESP), half of the participants received a quarter, and half did not.[5] Effectively, participants were randomly assigned to one of two conditions, a positive mood (quarter) and neutral mood (no quarter) control condition.

[5]The fairness of the die was evaluated in a set of independent trials. The die was rolled 102 times resulting in the following outcomes: 1 rolled 17 times, 2 rolled 15 times, 3 rolled 20 times, 4 rolled 20 times, 5 rolled 14 times, and 6 rolled 16 times. The chi-square test showed no evidence that any numbers rolled more or less often than chance, $X^2 = 1.88$, $df = 5$, *ns*.

Immediately following this procedure, participants responded to the same single-item measure of life satisfaction used in Studies 1–3. Participants in this study also reported their current mood and a question about how "lucky" they felt. In addition, they completed a domain satisfaction questionnaire (where they were asked about the degree to which they were satisfied with different domains of their life), the 5-item Satisfaction with Life Scale (SWLS; Diener et al., 1985), and a demographic questionnaire. After they completed all the questionnaires, participants were fully debriefed regarding the true purpose of the study.

Results

As can be seen in Table 8, participants who unexpectedly received a quarter did not report being in a more positive mood (d = −.03) or a less negative mood (d = .24) than participants in the no quarter control condition. In addition, results showed no difference in the single item of "lucky," with participants who won the quarter reporting the same levels of feeling lucky as participants who did not win the quarter (d = .25). Finally, receiving a quarter had no effect on life satisfaction, measured with a single item (d = −.09) or with the SWLS (d = .10). Thus, receiving a small amount of money did not affect participants' mood, feelings of being lucky, or their reports of global life satisfaction.

Local Meta-Analysis

We conducted a series of local meta-analyses to summarize the results of our studies along with the original findings in Table 1 (R script in the supplemental materials associated with this paper: https://osf.io/38bjg). We started by considering the overall effects of mood inductions on life satisfaction by combining the 9 effect sizes from our studies with the four effect sizes in Table 1. The overall effect size estimate for a random effects model using these data was d = .15 (95% CI = .04 to .27, p =.01); the overall estimate for a fixed effects model was d = .12 (95% CI = .04 to .21, p = .01). These are generally regarded as small effect sizes (at best) following standard conventions (e.g., Cohen, 1988). We adopted this approach for the happiness variables from Studies 1 to 8 and the three happiness effects sizes in Table 1. The overall effect size estimate using a random effects model was d = .27 (95% CI = [.02 to .52], p = .001); the estimate for a fixed effects model was d = .15 (95% CI = .06 to .24, p = .001). These are also considered a small effect sizes. Forest plots for the random effects models (including separate estimates for the original studies and our set of replications) are in Figures 1 and 2.

One concern with the overall analyses is that they collapse across different designs (e.g., writing manipulations versus video inductions) involving manipulations of different valances (neutral versus negative conditions, positive versus negative conditions). Thus, we conducted additional analyses restricted to studies with similar designs to provide finer grain perspectives on the effect size estimates. First, we combined the writing studies that used positive and negative conditions (Studies 1 to 3) with the results from Experiment 1 in Schwarz and Clore (1983). The meta-analytic effect sizes (using random effects models) were .27 (95% CI = −.01 to .54, p = .06) for life satisfaction and .59 (95% CI = −.25 to 1.42, p = .17) for happiness; neither effect was significant. Second, we combined the weather studies (our Studies 7 and 8 with Experiment 2 in Schwarz & Clore, 1983). The meta-

analytic effect sizes were .24 (95% CI = −.27 to .75, p = .36) for life satisfaction and .38 (95% CI = −.53 to 1.30, p = .41) for happiness. Again, neither average effect was significant. Last, to see whether our studies in which the mood induction manipulation checks showed that the mood inductions were successful resulted in larger effects, we combined all studies with successful mood inductions (Studies 1, 2, 3, 5, and 6). The meta-analytic effect sizes were .12 (95% CI = −.02 to .25, p = .10) for life satisfaction and .17 (95% CI = .05 to .29, p = .004) for happiness. Thus, the effect for happiness was significant but the effect for life satisfaction was not, and both estimated effects were very small.

In sum, the overall effect of mood manipulations on life satisfaction and happiness appears to be small rather than large. This conclusion holds looking across all studies and when focusing on specific comparisons. One complicating issue with the more specific analyses is that relatively few studies contribute to the analyses thereby increasing the standard errors. This contributed to larger p-values. To be sure, a visual inspection of the forest plots suggests that our effect size estimates were generally much smaller than those reported in the original studies.

Discussion

Global self-reports of subjective well-being are supposed to reflect a person's overall evaluation of the quality of his or her life as a whole (Diener et al., 1999). These reports should be largely independent of transitory moods induced by contextual features present in a given moment. Accordingly, the validity of measures of SWB would be called into question if subtle contextual factors have profound influences on self-reports (Schwarz & Strack, 1999). Prior studies have reported large context effects on SWB judgments. Given the influence that these studies have had on debates about the validity of global self-reports, especially in applied contexts (e.g., Kahneman, 1999; Kahneman & Krueger, 2006), systematic attempts at replication are needed.

The current article reports results of nine studies that were designed to evaluate how strongly experimental mood inductions affect respondents' subsequent reports of global SWB. Some of these nine studies (e.g., Studies 1 through 3, Studies 7 and 8) represent very close replications of prior research, whereas others (Studies 4, 5, 6, and 9) reflect partial or more conceptual replications of the early studies. Overall, the results of these studies suggest that mood effects on life satisfaction are typically small (at best) rather than large. In the small number of studies where mood effects on SWB judgments were statistically significant, they were substantially smaller than the effects found in the original study. In addition, many of the studies resulted in effects that were statistically indistinguishable from zero, even though our studies were powered to detect much smaller effects than those in the original studies.

As noted in the Introduction, there is no single criterion for determining whether a replication attempt obtains the same results as the original study, and indeed, conclusions about the extent to which our results replicate the original work vary somewhat depending on which criterion is used. We focused on (a) whether the study obtained a significant effect in the same direction as the original, (b) whether the original result falls outside of the

confidence interval of the replication attempt, and (c) whether the obtained results are too small to have been detected in the original study (Simonsohn, 2015b).

First, as can be seen in Figures 1 and 2, of the studies we conducted, only two of nine resulted in statistically significant effects when life satisfaction was the dependent variable, and only one of eight resulted in significant effects when global happiness was dependent variable. Looking at our second criterion, in none of the studies did the confidence interval for the effect size estimate overlap with the meta-analytic average of the effects from the four original studies. Finally, our third criterion was whether the effect found in our studies would have been detectable using Simonsohn's (2015b) "small telescopes" standard. The goal of this test is to determine whether the effect size found in the replication would have been detectable with 33% power in the original study. Given that the average per-cell sample size in the original studies was just 11, these studies could detect a d of .68 with 33% power. As can be seen in Figures 1 and 2, none of the effects found in our studies reached this threshold. Thus, using these three criteria, evidence for the replicability of the original effects is quite weak.

It is important to note, however, that there a number of reasons to be cautious in the interpretation of these results. First, as Figure 1 also shows, the meta-analytic effect size across all studies (including the originals) is significantly greater than zero, both for life satisfaction and happiness; and the effect size is significant for happiness (but not life satisfaction) even if the very large effects from the original studies are excluded. These results therefore suggest that the mood that a person is in at the time he or she makes a judgment can influence reports of SWB. However, the effects from our studies were substantially smaller than those from the original studies, with effect sizes that were, on average, less than one-tenth the size of the originals. One critical applied issue is whether the contextual effects are large enough to indicate that global measures of SWB lack validity for assessing the focal construct. As psychometricians have noted, experimental studies like those included in this paper, on their own, cannot address questions about validity (Alwin, 2007). Instead, additional studies that explicitly focus on the extent to which validity coefficients are affected by contextual features are needed. The small size of effects found in the current studies suggest that the effect on validity is likely not large; but additional research is needed to address this issue.

In addition, a second reason why it is important to interpret these results cautiously is that if we focus on the studies that are the closest and most straightforward replications of prior studies (Studies 1 through 3), the rate of "success" (as determined by significant effects) is two out of three rather than two out of nine for life satisfaction, and one out of three rather than one out of eight for happiness. However, even in these studies, the meta-analytic average (including the original effects) for life satisfaction (d = .27 [−.01, .54]) and happiness (d = .59 [−0.25, 1.42]) are not significantly different than zero. In addition, a meta-analysis of just the nine new studies that we conducted shows that the variability in effects is not significant either for life satisfaction, $Q(8) = 8.58$, or happiness, $Q(7) = 9.33$, *ns*, which suggests that the slightly larger effects found in Studies 1 and 3 may simply be the result of sampling error. Thus, even a selective focus on the studies that were closest to the

original findings would suggest that the typical effect of mood on life satisfaction judgments is considerably smaller than the effects reported in the original studies.

When considered in context of alternative approaches to understanding the links between mood and life satisfaction judgments, these results are actually quite consistent with a large body of research. For instance, if context effects were large, then short- and long-term stability of these measures would necessarily be quite low. Indeed, in their review, Schwarz and Strack (1999) explicitly stated that existing evidence suggesting that stability is weak provides support for their judgment model. However, their claim that "measures of SWB have low test-retest reliability, usually hovering around .40, and not exceeding .60 when the same question is asked twice during the same one-hour interview" (p. 62) is not supported by the large body of evidence that examines test-retest stability, even over very long periods of time. For instance, Schimmack and Oishi's (2005) meta-analytic results revealed that the aggregated test-retest reliability of global life satisfaction judgments is $r = .79$. This finding is consistent with other work estimating the reliability of single-item global measures of life satisfaction to be .73 across four large nationally representative panel studies (Lucas & Donnellan, 2012). Thus, the reliability of global well-being measures is comparable to other commonly used measures in psychology. In addition, about half of the reliable variance in life satisfaction tends to be stable even over very long periods of time (e.g., decades; Anusic & Schimmack, 2016; Lucas & Donnellan, 2012), casting further doubt on the proposition that contextual effects have a profound influence on global reports of SWB.

Furthermore, attempts to explicitly link naturally occurring variation in mood with changes in reports of life satisfaction have generally failed to find strong associations. For instance, Eid and Diener (2004) measured mood and global life satisfaction multiple times over the course of a semester, which allowed them to model stability and change in these constructs in a multistate-multitrait-multiconstruct model. Importantly, their study showed that the state component of global life satisfaction was quite small relative to the stable trait component (meaning that the scores did not vary much over time), and the within-person links between current mood and life satisfaction were relatively weak, showing that the two did not change together in substantial ways over time. Thus, more naturalistic investigations support our failure to find strong links between mood and self-reported life satisfaction.

Finally, other attempts to examine the effects of mood on life satisfaction judgments by examining the association between specific contextual factors that are thought to be related to mood and global well-being judgments have also found weak effects. For example, Lucas and Lawless (2013) used data from over one million respondents who had been surveyed throughout the U.S. on different days over a four-year period. They linked these responses to the weather on the day of the survey, and they found no effect on life satisfaction, even with an extremely high degree of power. As they review in their paper, strong and significant effects of weather on life satisfaction are typically quite difficult to find (also see Simonsohn, 2015a).

Potential Objections

In addition to issues related to judging the outcome of replication studies, there are other controversies about the explanation for failed replications. For example, small differences in

procedures—or even quite subtle differences in population, context, or even decade in which the study took place—are often proposed as potential explanations for failures to replicate. Of course, any failure to replicate an original finding could, in principle, be explained by such factors; the critical issue is whether such factors are plausible and how such factors place limits on the generalizability of particular effects.

One concern is that researchers who conduct replication studies do not go through the extensive pre-testing procedures that original researchers conduct to ensure that their manipulations and measures are appropriate in the population of interest. Fortunately, information about the extent of these procedures in the original study is available, as Schwarz and Clore provided details about the process that led to their results (Schwarz & Clore, 2003). In this follow-up paper written to celebrate the 20th anniversary of the original paper, they noted that instead of conducting extensive pre-tests for their original studies, they simply "[relied] on mental simulations of our own likely responses in setting up the procedures" (p. 298) and that the study worked on the very first try. Furthermore, subsequent studies showed that the basic design could be used in different populations (e.g., German students versus students from Illinois) and with substantial variation in procedures (e.g., Schwarz et al., 1987). Thus, this concern about calibrating the design to the population does not seem plausible in the current case as there is little reason to expect that the simple and straightforward designs could not be implemented with reasonable fidelity in our own studies. Indeed, no restrictions on the generalizability of these results were noted in the discussion of the original study.

A bigger potential concern has to do with how researchers should interpret null results from studies where manipulation checks fail to show that the manipulation worked (which was true of Studies 4, 7, 8, and 9). Indeed, even policies for "registered reports," where research papers are reviewed based solely on theory and design, often include provisions that manipulation checks must be successful for the paper to be published once results are known. Across the nine studies reported in this paper, only five resulted in significant differences in mood across the two mood conditions. This feature of our studies should certainly be taken into consideration when evaluating the implications of our results. However, there are three reasons why we think that all nine studies are important, despite the fact that mood manipulations only "worked" in five.

First, manipulation checks serve multiple purposes. Most importantly, they show whether the chosen manipulation has validity. In the case of Studies 1 through 4, for instance, a manipulation check can help determine whether the act of writing about positive and negative life events affects self-reports of mood. Although Schwarz and Clore (1983) already demonstrated the validity of this particular induction, it is still useful to show that the manipulation works in a different population with slightly modified procedures, as we did in these studies.

If a manipulation with demonstrated validity does not result in statistically significant result when manipulation checks are conducted, however, it is unclear how one should interpret the results. It is possible, for instance, that the manipulation did have the intended effect, but that because of sampling error, the manipulation check did not reach the standard threshold for

statistical significance (i.e., the outcome was a false negative for the manipulation check). Indeed, even for well-validated manipulations, and even in studies where the primary predicted effect was obtained, we should expect some manipulation checks to fail solely due to sampling error.

Second, and more importantly, because researchers have flexibility in the analyses they conduct and report, manipulation checks allow for increased "researcher degrees of freedom" regarding the findings that make it into the published literature. If, for instance, a researcher finds a predicted effect with a previously successfully experimental paradigm, that researcher may still report the finding as a "hit" even if the manipulation check failed, drawing on the apparent face validity of the manipulation or the precedence in the literature. In contrast, that same researcher may decide not to publish the finding if the failed manipulation check occurs alongside an unpredicted result; the failed manipulation check serves as a post hoc explanation of the unpredicted result. Indeed, examples of this can be found within the literature that is the focus of this paper. For instance, in both Studies 1 and 3 in Strack et al. (1985), a study that used the same procedures as Studies 1 through 3 of this paper, manipulation checks did not result in significant mood differences across the critical conditions (though there was a significant interaction for one of these two studies), yet the authors still interpreted the results because they were in line with predictions. Our view is that the aggregate body of research benefits from the reporting of all results, so we reported all studies in this package. It is important to reiterate, however, that our bottom line conclusions would be the same if we focus only on studies where mood manipulation checks showed significant differences.

Each of the previous two points rests on the assumption that well-validated experimental paradigms are being used in the studies. However, a third reason for publishing null results even when manipulation checks show that the manipulation failed relates to concerns about the validity of the manipulation itself. In short, when the validity of the manipulation used in an original study is in question, nonsignificant manipulation checks in a replication study cast doubt on the validity of that manipulation, which can lead to questions about the plausibility of results reported in the original study.

For instance, in Study 2 of Schwarz and Clore (1983), the authors assumed that weather—or at least unusually pleasant weather—would have strong average effects on mood, which should then strongly influence life satisfaction judgments. As they later noted, they did not do any pre-testing to confirm this assumption, and they did not cite a large body of existing literature that demonstrated the efficacy of weather for this purpose; they simply relied on their intuitions about weather and its effects (Schwarz & Clore, 2003). However, a close look at the current literature shows that weather effects on mood are actually quite difficult to find, so there is very little empirical evidence that supports the intuitions that guided the initial choice of procedures. Although studies that report positive results for the effect of weather on mood do exist, the specific findings from these publications rarely replicate, as Lucas and Lawless (2013) noted. Instead, most studies report distinct weather/mood associations that rarely emerge in identical ways across studies (a pattern that leads to concerns about flexibility in data analysis and publication bias). Furthermore, the largest studies that have examined the links between weather and mood typically find extremely

small effects, if the effects are even detectable (e.g., Denisson, Butalid, Penke, & Van Aken, 2008; Klimsta, Frijns, Keijsers, Denisson, Raaijmakers, Quinten, van Aken, Koot, van Lier, & Meeus, 2011; Koots, Realo, & Allik, 2011; Watson, 2000; though at least some of these studies suggest that people vary systematically in their mood response to weather, an effect that neither our studies nor the original by Schwarz and Clore were designed to detect). Thus, there are strong reasons to doubt the plausibility of weather as valid mood induction, especially of the size reported in Schwarz and Clore (1983).

Relatedly, although the strongest concerns about failed manipulation checks can be directed at our Study 9 (indeed, we believe that this is the weakest study in this package), we hope this study challenges readers to carefully examine past work that has used similar mood induction techniques. As noted in the Introduction, this study was inspired by Schwarz (1987), where he found that finding a dime is associated with higher levels of life satisfaction. Although this effect of finding the dime on global measures of well-being was not actually significant in Schwarz's study, the study is still usually interpreted as a positive result for the general idea that mood influences satisfaction judgments. We tried a number of pilot studies to mimic the experience of finding a trivially small amount of money but were unsuccessful at identifying conditions that reliably impacted mood. Future studies should try alternative designs to try to replicate this particular effect.

Finally, it is worth pointing out that even when we restrict conclusions to the five studies where the mood inductions were successful (or even to the three most direct replications where mood inductions were successful), our conclusions would be the same: The effect of mood on life satisfaction judgments is, at best, very small in size. Thus, even when the most restrictive criteria are used, these studies challenge Schwarz and Strack’s (1999) conclusion that mood effects on global well-being judgments are large. In turn, this leads to questions about the practical importance of these effects, questions that will need to be followed up with additional studies that use different designs to explicitly test the implications for validity.

As a final limitation, we acknowledge that our studies were not pre-registered, meaning that we did not do everything in our power to avoid our own researcher degrees of freedom. We should note that this concern is mitigated somewhat by the fact that our replications followed the original designs as closely as possible, which meant that there were few decisions that had to be made regarding which particular outcomes to include or which specific analyses to conduct. In addition, we have made all materials and data available in the supplemental materials, so interested readers can determine whether there are any other measures that could have been included in our analyses. Finally, we have conducted no other studies that are relevant to these hypotheses.

Implications and Conclusions

Taken together, the results of these studies have important implications for researchers and policy makers interested in measuring subjective quality of life. In the past, concerns have been raised about the extent to which minor contextual factors can influence what is supposed to be a relatively stable judgment. A small number of widely cited studies have suggested that such context effects can be “profound”. However, the studies that have found

strong context effects tend to use small sample sizes, and few direct replications have been reported in the literature. The current work adds to the literature by presenting a series of large-sample direct and conceptual replications of the previously reported context effects. Although these do not represent the final word on mood effects on global well-being judgments, they suggest that more work is needed before we accept the robustness of the original effects as indicators of severe flaws in global self-reports of SWB.

As the need for cost-effective, reliable methods to accurately assess societal quality of life grows, evidence in favor of the continued use of global measures of life satisfaction such as the ones evaluated in the current studies has tremendous importance for research and policy. To be sure, the results of this study do not imply that global self-report measures like the ones evaluated in these studies are necessarily the optimal method of assessing quality of life, and we firmly support the notion that research must continue to refine and improve how researchers assess quality of life and achieve balance between measurement accuracy and ease of assessment. This is vital to the understanding of SWB on a societal level and to the application of insights and knowledge garnered from research using these global measures. However, the results reported here suggest that the current mood of respondents, which research has shown can vary greatly due to subtle changes in an individual's context, does not appear to have an especially large impact on respondents' reports of global well-being. This finding contrasts with a well-known critique of global SWB measures, which suggests that positive and negative feelings due to mood can have a "profound" impact one's overall judgment of quality of life. Our results suggest that people's life satisfaction judgments may be less influenced by surrounding context than previously believed.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This research was supported by NIA grant AG040715 awarded to Brent Donnellan and Richard Lucas.

References

Alwin, DF. Margins of error: A study of reliability in survey measurement. Vol. 547. John Wiley & Sons; 2007.

Anusic I, Lucas RE, Donnellan MB. The Validity of the Day Reconstruction Method in the German Socio-Economic Panel Study. Social Indicators Research. (in press).

Anusic I, Schimmack U. Stability and change of personality traits, self-esteem, and well-being: Introducing a Meta-Analytic Stability and Change (MASC) model of retest correlations. Journal of Personality and Social Psychology. (in press).

Baumeister RF, Bratslavsky E, Finkenauer C, Vohs KD. Bad is stronger than good. Review of General Psychology. 2001; 5:323–370.

Bradburn, NM. The structure of psychological well-being. Chicago: Aldine; 1969.

Deaton, A. The Financial Crisis and the Well-Being of Americans. Cambridge, MA: National Bureau of Economic Research (NBER); 2011.

Denissen JJ, Butalid L, Penke L, Van Aken MA. The effects of weather on daily mood: A multilevel approach. Emotion. 2008; 8:662–667. [PubMed: 18837616]

Diener E. Subjective well-being. Psychological Bulletin. 1984; 95:542–575. [PubMed: 6399758]

Diener E, Emmons RA, Larsen RJ, Griffin S. The Satisfaction with Life Scale. Journal of Personality Assessment. 1985; 49:71–75. [PubMed: 16367493]

Diener, E., Lucas, R., Schimmack, U., Helliwell, J. Well-being for public policy. New York, NY: Oxford University Press; 2009.

Diener E, Suh EM, Lucas RE, Smith HL. Subjective well-being: Three decades of progress. Psychological Bulletin. 1999; 125:276–302.

Goldberg LR, Johnson JA, Eber HW, Hogan R, Ashton MC, Cloninger CR, Gough HG. The international personality item pool and the future of public-domain personality measures. Journal of Research in Personality. 2006; 40:84–96.

Gross JJ, Levenson RW. Emotion elicitation using films. Cognition and Emotion. 1995; 9:87–108.

Gross JJ, Sutton SK, Ketelaar T. Relations between affect and personality: Support for affect-level and affective-reactivity views. Personality and Social Psychology Bulletin. 1998; 24:279–291.

Ioannidis JPA. Why most published research findings are false. PLoS Med. 2005; 2:e124. [PubMed: 16060722]

Isen AM, Levin PF. Effect of feeling good on helping: cookies and kindness. Journal of personality and social psychology. 1972; 21(3):384–388. [PubMed: 5060754]

Isen AM, Simmonds SF. The effect of feeling good on a helping task that is incompatible with good mood. Social Psychology. 1978:346–349.

Kahneman, D. Objective happiness. In: Kahneman, D.Diener, E., Schwarz, N., editors. Well-being: Foundations of hedonic psychology. New York: Russell Sage Foundation Press; 1999. p. 3-25.

Kahneman D, Krueger AB. Developments in the measurement of subjective well-being. The Journal of Economic Perspectives. 2006; 20:3–24.

Klimstra TA, Frijns T, Keijsers L, Denissen JJ, Raaijmakers QA, van Aken MA, Meeus WH. Come rain or come shine: Individual differences in how weather affects mood. Emotion. 2011; 11:1495–1499. [PubMed: 21842988]

Kööts L, Realo A, Allik J. The influence of the weather on affective experience. Journal of Individual Differences. 2011; 32:74–84.

Lucas RE, Donnellan MB. Estimating the reliability of single-item life satisfaction measures: Results from four national panel studies. Social Indicators Research. 2012; 105:323–331. [PubMed: 23087538]

Lucas RE, Lawless NM. Does life seem better on a sunny day? Examining the association between daily weather conditions and life satisfaction judgments. Journal of Personality and Social Psychology. 2013; 104:872–884. [PubMed: 23607534]

Lyubomirsky S, Lepper H. A measure of subjective happiness: Preliminary reliability and construct validation. Social Indicators Research. 1999; 46:137–155.

Maxwell SE, Lau MY, Howard GS. Is psychology suffering from a replication crisis? What does “failure to replicate” really mean? American Psychologist. 2015; 70:487–498. [PubMed: 26348332]

Schwarz, N. Stimmung als Information: Untersuchungen zum Einfluβ von Stimmungen auf die Bewertung des eigenen Lebens (Mood as information on the impact of moods on evaluations of one’s life). Heidelberg: Springer-Verlag; 1987.

Schwarz N, Clore GL. Mood, misattribution, and judgments of well-being: Informative and directive functions of affective states. Journal of Personality and Social Psychology. 1983; 45:513–523.

Schwarz N, Clore GL. Mood as information: 20 years later. Psychological Inquiry. 2003; 14:296–303.

Schwarz, N., Strack, F. Reports of subjective well-being: Judgmental processes and their methodological implications. In: Kahneman, D.Diener, E., Schwarz, N., editors. Well-being: The foundations of hedonic psychology. New York: Russell Sage Foundation; 1999. p. 61-84.

Schwarz N, Strack F, Mai HP. Assimilation and contrast effects in part-whole question sequences: A conversational logic analysis. Public Opinion Quarterly. 1991; 55(1):3–23.

Schwarz N, Strack F, Kommer D, Wagner D. Soccer, rooms, and the quality of your life: Mood effects on judgments of satisfaction with life in general and with specific domains. European Journal of Social Psychology. 1987; 17(1):69–79.

Schimmack U, Oishi S. The influence of chronically and temporarily accessible information on life satisfaction judgments. Journal of Personality and Social Psychology. 2005; 89(3):395–406. [PubMed: 16248721]

Simonsohn U. Rain & Happiness: Why Didn't Schwarz & Clore (1983) 'Replicate' ? 2015a Nov 16. [Retrieved November 24, 2015] from http://datacolada.org/2015/11/16/43_rain_and_happiness/.

Simonsohn U. Small Telescopes Detectability and the Evaluation of Replication Results. Psychological Science. 2015b; 26:559–569. [PubMed: 25800521]

Stiglitz JE, Sen A, Fitoussi J. Report by the commission on the measurement of economic performance and social progress. Commission on the Measurement of Economic Performance and Social Progress. 2009 Retrieved from http://www.stiglitz-sen-fitoussi.fr/documents/rapport_anglais.pdf.

Strack F, Martin LL, Schwarz N. Priming and communication: Social determinants of information use in judgments of life satisfaction. European Journal of Social Psychology. 1988; 18(5):429–442.

Strack F, Schwarz N, Gschneidinger E. Happiness and reminiscing: The role of time perspective, mood, and mode of thinking. Journal of Personality and Social Psychology. 1985; 49:1460–1469.

University of Waterloo. CIW releases Canada’s first-ever index to measure national wellbeing. 2011 Retrieved from http://uwaterloo.ca/applied-health-sciences/news/ciw-releases-wellbeing-index.

U. S. Department of Health and Human Services. Healthy People 2020. 2014. Retrieved from http://www.healthypeople.gov/2020/default.aspx

Watson, D. Mood and temperament. New York, NY: Guilford Press; 2000.

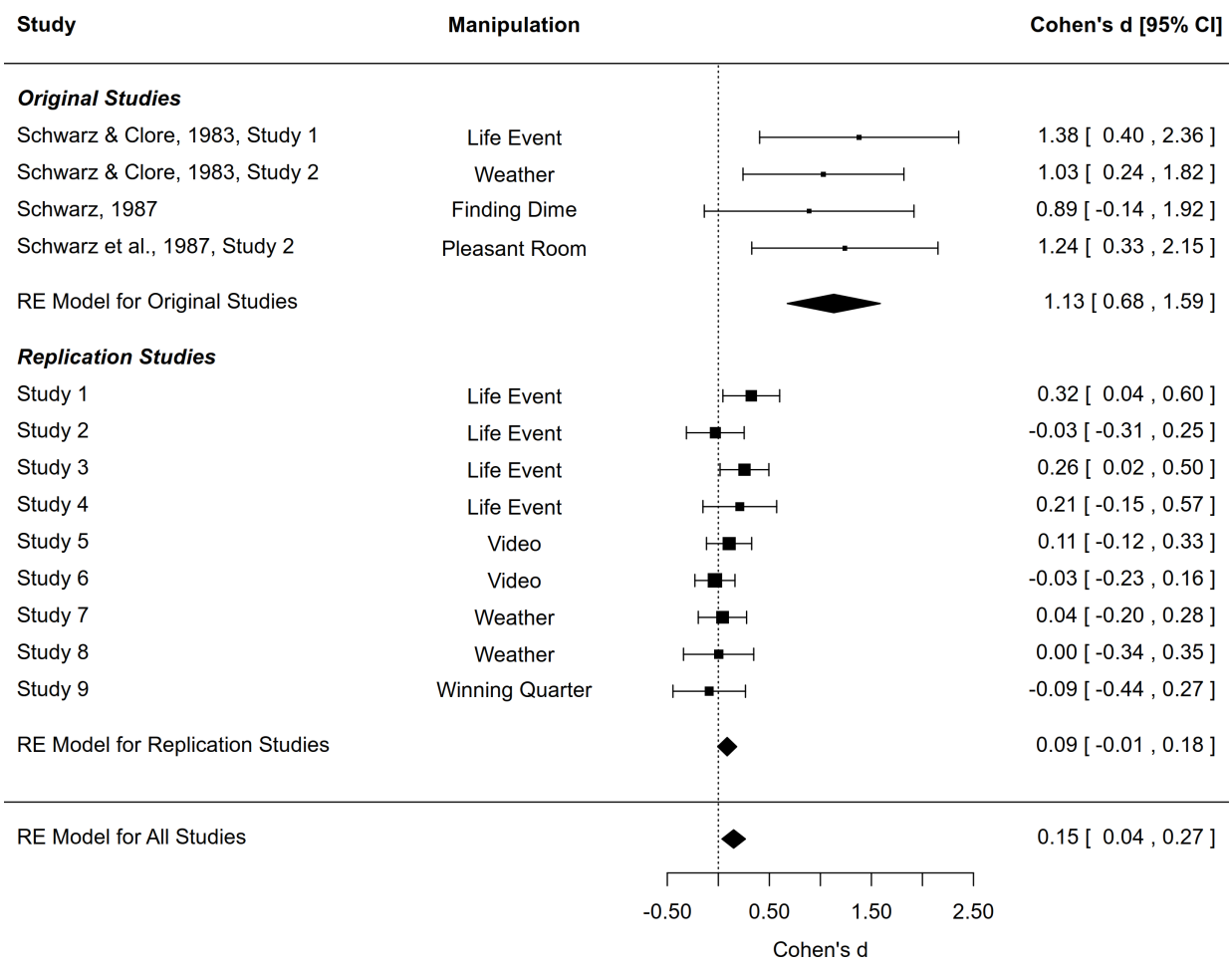

Figure 1.
Forest plot of results for life satisfaction, with meta-analytic summary.

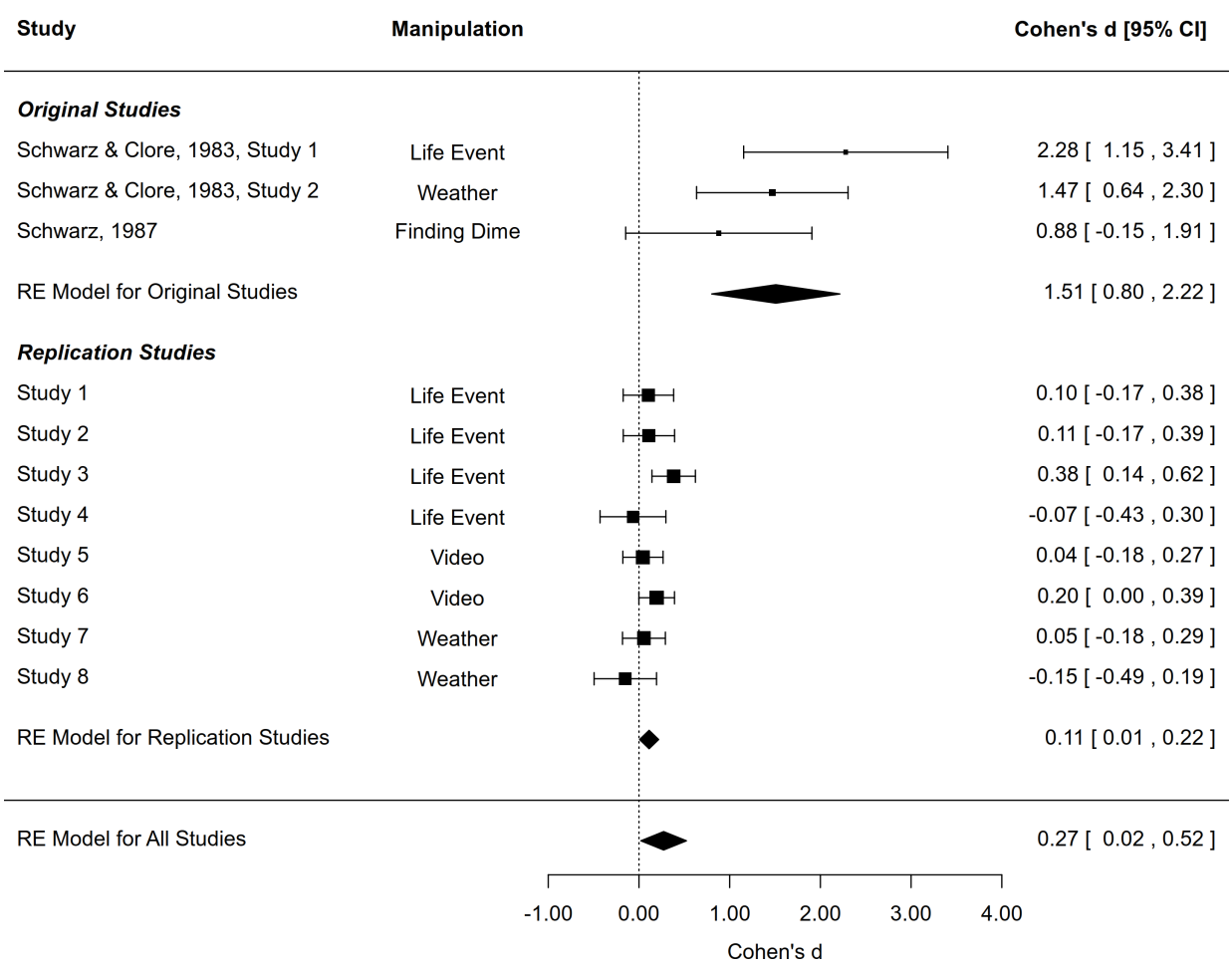

Figure 2.
Forest plot of results for happiness, with meta-analytic summary.

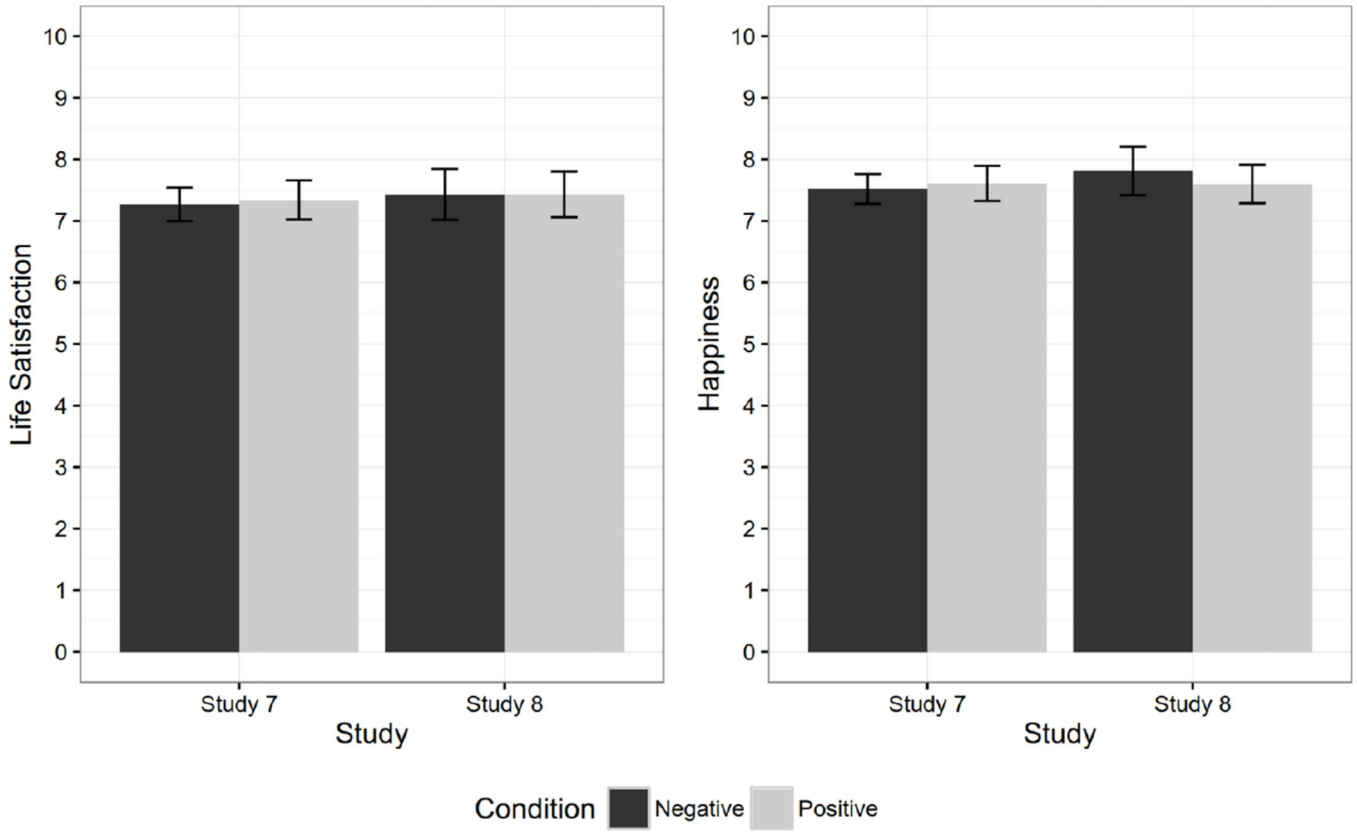

Figure 3.
Means for life satisfaction (left panel) and happiness (right panel), with 95% confidence intervals, for the pleasant and unpleasant weather conditions in Studies 7 and 8.

Table 1

Results of Past Research

Study Name	Type of Mood Manipulation	N	Cohen's *d* Life Satisfaction	Cohen's *d* Happiness
Schwarz & Clore, 1983: Experiment 1	Writing: Participants are asked to describe positive or negative life events	61	1.38	2.28
Schwarz & Clore, 1983: Experiment 2	Weather: Participants are asked about their life satisfaction on a sunny or a rainy day	84	1.03	1.47
Strack, Schwarz, & Gschneidinger, 1985: Experiment 1	Writing: Participants are asked to describe positive or negative life events in the present or the past	51		
Strack, Schwarz, & Gschneidinger, 1985: Experiment 2	Writing: Participants are asked to describe positive or negative life events using detailed or short descriptions	36		
Strack, Schwarz, & Gschneidinger, 1985: Experiment 3	Writing: Participants are asked to describe positive or negative life events and why or how it happened	64		
Schwarz, 1987	Dime: Participants in the positive mood condition found a dime that had been surreptitiously left for them	16	.89	.88
Schwarz et al., 1987: Experiment 1	Soccer Game: Participants are asked about their life satisfaction before or after their soccer team won or lost a game	55		
Schwarz et al., 1987: Experiment 2	Room: Participants are asked about their life in a pleasant or unpleasant room	22	1.24	

Note. Empty cells reflect studies where not enough information was provided to calculate effect sizes. Effect sizes calculations are detailed in the on-line supplement.

Table 2

Participant characteristics across the studies.

	Gender						Age			
	N Negative	**N Positive**	**Female**	**Male**	**Other**	**Missing**	***M***	***SD***	**Range**	**Semester Completed**
Study 1: Writing 1	111	91	152	49	0	1	19.49	2.5	18–42	Fall 2012
Study 2: Writing 2	119	81	151	44	1	4	19.33	1.98	18–38	Fall 2012
Study 3: Writing 3	136	133	194	66	0	9	19.30	2.54	18–40	Spring 2013
Study 4: Writing 4	61	57*	80	36	0	2	18.90	1.02	18–22	Fall 2015
Study 5: Video 1	155	165	222	91	1	6	19.62	1.68	18–34	Fall 2013
Study 6: Video 2	211	190	330	70	1	0	19.27	1.38	18–25	Spring 2014
Study 7: Weather 1	253	208	347	103	2	9	19.8	1.8	18–35	Spring 2013
Study 8: Weather 2	183	192	287	87	0	1	19.5	1.4	18–28	Spring 2014
Study 9: Quarter	58	64	92	30	0	0	19.79	3.32	18–49	Spring 2012

Note.

* = neutral condition rather than positive. Reported data for Studies 7 and 8 are from the initial baseline survey.

Table 3

Means and tests of differences in means across the writing conditions in Studies 1 through 3.

	N		*M (SD)*					
	Positive	**Negative**	**Positive**	**Negative**	***t***	***df***	***p***	***d* (95% CI)**
Study 1 Writing 1								
Positive Mood	90	111	3.48 (0.83)	2.87 (0.93)	4.90	199	<.001	0.69 (0.41, 0.98)
Negative Mood	90	111	1.53 (0.68)	2.16 (0.99)	−5.10	199	<.001	−0.72 (−1.01, −0.44)
Happy Life	90	111	5.51 (0.94)	5.41 (0.92)	0.73	199	.46	0.10 (−0.17, 0.39)
Life Satisfaction	90	111	7.40 (1.63)	6.85 (1.78)	2.28	199	.02	0.32 (0.04, 0.60)
Happy (Now)	89	111	5.12 (1.14)	4.50 (1.26)	3.65	198	<.001	0.52 (0.23, 0.79)
Unhappy (Now)	90	111	2.32 (1.29)	3.15 (1.46)	−4.23	199	<.001	−0.60 (−0.88, −0.31)
Good	90	111	5.21 (1.28)	4.51 (1.37)	3.70	199	<.001	0.53 (0.26, 0.82)
Bad	90	111	2.18 (1.26)	2.95 (1.54)	−3.81	199	<.001	−0.54 (−0.81, −.24)
Study 2 Writing 2								
Positive Mood	81	118	3.09 (0.94)	2.74 (1.04)	2.41	197	.02	0.34 (0.06, 0.63)
Negative Mood	81	118	1.93 (0.94)	2.15 (1.03)	−1.50	197	.14	−0.21 (−0.50, 0.07)
Happy Life	81	118	5.40 (1.06)	5.27 (1.15)	0.77	197	.44	0.11 (−0.17, 0.40)
Life Satisfaction	81	118	6.99 (1.78)	7.04 (1.84)	−0.21	197	.83	−0.03 (−0.31, 0.26)
Happy (Now)	81	118	4.68 (1.37)	4.37 (1.33)	1.58	197	.12	0.23 (−0.05, 0.51)
Unhappy (Now)	81	118	3.02 (1.60)	3.32 (1.54)	−1.32	197	.19	−0.19 (−0.47, 0.09)
Good	81	117	4.73 (1.31)	4.55 (1.43)	0.91	196	.37	0.13 (−0.15, 0.41)
Bad	81	118	2.69 (1.51)	3.08 (1.60)	−1.71	197	.09	−0.24 (−0.53, 0.04)
Study 3 Writing 3								
Positive Mood	132	136	3.42 (0.96)	2.83 (1.05)	4.81	266	<.001	0.59 (0.36, 0.85)
Negative Mood	132	136	1.53 (0.63)	1.97 (0.89)	−4.70	266	<.001	−0.58 (−0.83, −0.34)
Happy Life	132	136	5.61 (1.02)	5.20 (1.15)	3.12	266	.00	0.38 (0.13, 0.62)
Life Satisfaction	133	136	7.29 (1.87)	6.79 (2.09)	2.09	267	.04	0.26 (0.01, 0.49)
Happy (Now)	133	136	5.15 (1.41)	4.38 (1.44)	4.42	267	<.001	0.54 (0.30, 0.78)
Unhappy (Now)	133	136	2.38 (1.39)	3.24 (1.50)	−4.87	267	<.001	−0.60 (−0.81, −0.35)
Good	133	136	5.29 (1.32)	4.65 (1.46)	3.80	267	<.001	0.47 (0.22, 0.70)
Bad	133	136	2.18 (1.36)	2.98 (1.51)	−4.54	267	<.001	−0.56 (−0.80, −0.31)

Table 4

Means and tests of differences in means across the writing conditions in Study 4.

	Negative N	Neutral N	Negative M	Neutral M	Negative SD	Neutral SD	t	p	df	d[95%CI]
Life Satisfaction	61	57	5.17	5.43	1.42	0.99	1.10	0.26	116	0.21[−0.15, 0.57]
Subjective Happiness	61	56	5.12	5.04	1.29	1.11	−0.35	0.73	115	−0.07[−0.43, 0.30]
Positive Affect	61	57	4.07	3.93	1.18	1.16	−0.63	0.53	116	−0.12[−.25, 0.48]
Negative Affect	61	57	2.23	1.89	1.48	1.48	−1.20	0.22	116	−0.22[−0.59, 0.14]
Happy/Unhappy	61	57	4.74	4.89	1.29	1.06	0.72	0.47	116	0.13[−0.23, 0.49]
Good/Bad	61	57	4.97	5.21	1.34	1.11	1.07	0.29	116	0.20[−0.17, .0.56]

Table 5

Means and tests of differences in means across the video conditions in Studies 5 and 6

	N		*M (SD)*					
	Positive	**Negative**	**Positive**	**Negative**	***t***	***df***	***p***	***d (95% CI)***
Study 5 Video 1								
Positive Mood	162	152	3.51 (0.86)	3.23 (0.97)	2.64	312	0.009	0.30 (0.08, 0.53)
Negative Mood	162	152	1.63 (0.76)	1.88 (0.82)	−2.83	312	0.004	−0.32 (−0.55, −0.10)
Happy Life	161	152	5.55 (0.89)	5.51 (0.93)	0.39	311	0.70	0.04 (−0.18, 0.27)
Life Satisfaction	162	152	7.35 (1.76)	7.16 (1.81)	0.93	312	0.35	0.11 (−0.12, 0.33)
SWLS	161	152	5.07 (1.15)	4.88 (1.16)	1.48	312	0.14	0.17 (−0.10, 0.35)
Study 6 Video 2								
Positive Mood	190	211	3.61 (0.88)	3.06 (0.98)	5.90	399	<.001	0.59 (0.39, 0.79)
Negative Mood	190	211	1.64 (0.84)	2.10 (0.95)	−5.03	399	<.001	−0.50 (−0.71, −0.31)
Happy Life	189	211	5.55 (1.03)	5.34 (1.1)	1.95	398	0.052	0.20 (0.00, 0.39)
Life Satisfaction	189	211	7.05 (1.89)	7.11 (1.77)	−0.33	398	0.74	−0.03 (−0.23, 0.16)
SWLS	189	210	4.99 (1.10)	4.95 (1.77)	0.36	397	0.72	0.04 (−0.16, 0.23)

Table 6

Weather summary for the target and preceding days.

	Temperature (max)	Temperature (mean)	Temperature (min)	Precipitation (in)	Cloud cover	Date
Study 7						
Good weather 1						
Mean over 5 previous days	46	35	24	0.00	4	March 30 – April 3, 2013
Target day	56	42	28	0.00	1	April 4, 2013
Good weather 2						
Mean over 5 previous days	47	41	35	0.34	7	April 11–15, 2013
Target day	56	50	43	0.30	6	April 16, 2013
Bad weather 1						
Mean over 5 previous days	33	28	23	0.08	7	March 15–19, 2013
Target day	25	20	14	0.00	6	March 20, 2013
Bad weather 2						
Mean over 5 previous days	54	44	32	0.35	5	April 5–9, 2013
Target day	41	39	36	0.51	8	April 10, 2013
Study 8						
Good weather						
Mean over 5 previous days	42	33	24	0.05	5	March 26–30, 2014
Target day	60	46	31	0.00	0	March 31, 2014
Bad weather						
Mean over 5 previous days	50	40	28	0.01	2	March 29–April 2, 2014
Target day	41	37	33	0.04	8	April 3, 2014
Means for target days (*SD*)						
Good weather	57 (2)	46 (4)	34 (8)	0.10 (0.17)	2 (3)	
Bad weather	36 (9)	32 (10)	28 (12)	0.18 (0.28)	7 (1)	

Notes: Temperature reported in degrees Fahrenheit. Cloud cover is rated on a scale from 0 (no cloud cover) to 8 (full cloud cover).
Source: http://www.wunderground.com.

Table 7

Means and tests of differences in means across the weather conditions in Studies 7 and 8.

	N		*M (SD)*					
	Good	**Bad**	**Good**	**Bad**	***d***	***df***	***p***	***d* (95% CI)**
Study 7 (Spring 2013)								
Life happy	117	167	7.61 (1.54)	7.52 (1.55)	0.46	282	.64	0.05 (−0.29, 0.18)
Life change	117	167	6.08 (2.42)	5.83 (2.58)	0.83	282	.41	0.10 (−0.33, 0.13)
Life satisfaction	116	168	7.34 (1.71)	7.27 (1.76)	0.34	282	.74	0.04 (−0.27, 0.19)
Mood PA	117	167	3.44 (0.88)	3.28 (0.90)	1.44	282	.15	0.17 (−0.41, 0.06)
Mood NA	117	167	1.97 (0.92)	1.88 (0.89)	0.85	282	.40	0.10 (−0.34, 0.13)
DRM PA	78	113	2.95 (1.09)	2.65 (1.20)	1.75	189	.08	0.26 (−0.03, 0.55)
DRM NA	78	113	1.21 (0.77)	1.39 (0.98)	−1.35	189	.18	−0.20 (−0.49, 0.09)
Day satisfaction	81	116	6.68 (2.00)	6.51 (2.02)	0.59	195	.56	0.08 (−0.20, 0.37)
Study 8 (Spring 2014)								
Life happy	67	63	7.60 (1.28)	7.81 (1.56)	−0.85	128	.40	0.15 (−0.19, 0.51)
Life change	67	63	5.87 (2.42)	6.41 (1.85)	1.44	128	.15	0.25 (−0.10, 0.60)
Life satisfaction	67	63	7.43 (1.52)	7.43 (1.64)	0.02	128	.99	−0.00 (−0.35, 0.35)
Mood PA	67	63	3.48 (0.78)	3.55 (0.75)	−0.50	128	.61	0.09 (−0.27, 0.43)
Mood NA	67	63	1.87 (0.80)	1.87 (0.80)	−0.05	128	.96	0.01 (−0.33, 0.37)
DRM PA	110	90	2.98 (1.07)	2.83 (1.02)	1.02	198	.31	0.15 (−0.14, 0.42)
DRM NA	110	90	1.58 (1.01)	1.33 (0.82)	1.89	198	.06	0.27 (−0.03, 0.54)
Day satisfaction	111	92	6.50 (2.06)	6.92 (1.71)	−1.56	201	.12	−0.22 (−0.49, 0.07)

Table 8

Means and tests of differences in means across the two groups in Study 9.

	N		*M (SD)*					
	Win	**Lose**	**Win**	**Lose**	***t***	***df***	***p***	***d (95% CI)***
Positive Mood	64	58	3.36 (0.90)	3.39 (0.81)	−0.17	120	0.87	−0.03 (−0.39, 0.32)
Negative Mood	64	58	1.78 (0.68)	1.61 (0.72)	1.32	120	0.19	0.24 (−0.12, 0.60)
Life Satisfaction	64	58	7.42 (1.40)	7.55 (1.54)	−0.49	120	0.63	−0.09 (−0.44, 0.27)
SWLS	64	58	5.09 (1.19)	4.97 (1.18)	0.55	120	0.58	0.10 (−0.25, 0.46)
Lucky	63	57	2.79 (1.22)	2.49 (1.26)	1.34	118	0.18	0.25 (−0.11, 0.61)