University of St. Thomas, Minnesota
UST Research Online

Marketing Faculty Publications | Marketing

2018

The Self-Contaminating Nature of Repeated Reports of Negative Emotions

Omesh Johar
Central State University, ojohar@centralstate.edu

Aaron M. Sackett
University of St. Thomas, Minnesota, sackett@stthomas.edu

Follow this and additional works at: https://ir.stthomas.edu/ocbmktgpub

 Part of the Marketing Commons

This Article is brought to you for free and open access by the Marketing at UST Research Online. It has been accepted for inclusion in Marketing Faculty Publications by an authorized administrator of UST Research Online. For more information, please contact asle4660@stthomas.edu.

BASIC AND APPLIED SOCIAL PSYCHOLOGY
2018, VOL. 40, NO. 5, 293–307
https://doi.org/10.1080/01973533.2018.1496336

Routledge
Taylor & Francis Group

Check for updates

The Self-Contaminating Nature of Repeated Reports of Negative Emotions

Omesh Johar[a] and Aaron M. Sackett[b]

[a]Central State University; [b]University of St. Thomas

ABSTRACT

Three experiments examined effects of measuring self-reported emotional intensity on subsequent self-reported emotional intensity. Across 3 experiments, we induced sadness, envy, and happiness and manipulated the number of emotional intensity measurements. In all experiments, repeated measurement led to weaker intensity of negative emotions than did a single measurement. Although the intensity of happiness was unaffected by repeated measurement, data suggest that measurements interfered with ongoing emotional experience. We suggest that our findings have methodological, conceptual, and practical implications, but perhaps foremost is the warning that social scientists may have greater cause for caution regarding repeated self-report measures than previously thought.

Psychological measurements are imperfect, and researchers understand that measurement error is inevitable (Cole & Preacher, 2014; Schmidt & Hunter, 1996). With this understanding, it is convenient to assume that measurement error is an artifact of observation, but the signal or the true value of a variable remains unaffected by the measurement process. This assumption is widespread in psychological science (for exceptions, see Knowles, Coker, Scott, Cook, & Neville, 1996; Sharpe & Gilbert, 1998; Shrout et al., 2018). If pressed, many researchers would acknowledge the possibility of measurement interference and would perhaps cite the Hawthorne effect (Landsberger, 1958), which describes behavioral changes stemming from participants' awareness that they are being observed. However, most current empirical work ignores the possibility that measures might affect the psychological constructs being observed. The goal of the current research was to examine the effect of measurement on the psychological phenomena under investigation.

To highlight our concerns regarding psychological measurements, let us use an example from the physical sciences in which the term *observer effect* is commonly used to refer to disturbances in a system due to measurement (e.g., Buks, Schuster, Heiblum, Mahalu, & Umansky, 1998; Riley & Steitz, 2013): When one measures the temperature of hot water with a glass thermometer, some heat is lost from the water to the thermometer. As a result, temperature measurements are likely biased. With a large vessel such as a barrel, the bias might be negligible because only a tiny fraction of the water's heat is transferred to the thermometer. However, if one uses the same thermometer to measure temperature in a small vessel, such as a 50-ml beaker, the heat lost to the thermometer may significantly threaten the validity of the observation. Furthermore, subsequent measurements of the water's temperature will yield different results than if the water had not been previously measured. Thus, if one wishes to measure natural temperature change over time, one must take into account the self-contaminating nature of the measurement process, whereby heat lost during each measurement could change the temperature being measured.

Similar concerns about self-contamination may apply to measurements of psychological phenomena such as emotions. If, like physical heat, some amount of emotional "heat" were lost due to measurement, it could threaten the validity of that observation and especially any subsequent observations. Although psychologists are already keenly aware that any psychological measure provides only an imperfect estimate of the value of a psychological construct, we propose a potentially more threatening possibility: that such measures might also inadvertently change the value of the constructs they are intended to observe. In our attempt to uncover observer effects in psychology, we focus the present investigation on the measurement of emotions through self-reports. From this point forward, we use the

CONTACT Omesh Johar ojohar@centralstate.edu Central State University, 1400 Brush Row Road, Wilberforce, OH 45384, USA.
© 2018 Taylor & Francis Group, LLC

truncated term *repeated measurement* to refer specifically to repeated measurement via self-report.

A brief overview of observer effects in social science

As just mentioned, there exists a limited but notable collection of past research that has examined the possibility of observer effects in psychological or behavioral measures. The arguably best-known example of observation interfering with the psychological phenomena being measured is the Hawthorne effect (Landsberger, 1958). Although the Hawthorne effect has received criticism for weak evidence in the original data and multiple researchers' struggles to replicate findings (Carey, 1967; Levitt & List, 2011; Miller & Form, 1951), it introduced many social and psychological scientists to the possibility of observer effects. At the very least, psychologists have since acknowledged that measurement can create "demand characteristics," which is a broad term used to describe the effects of social or procedural cues that participants may pick up on during a study and of participants' motivations related to being observed, including attempts to be a "good subject" (Orne, 1962; see also Adair, 1984).

Whereas the Hawthorne effect drew attention to the effects of simply measuring a construct, the effects of repeatedly measuring the same construct were pointed out by research on attitudes. It has been shown that repeated measurements or expressions of attitudes can increase attitude accessibility (Powell & Fazio, 1984), attitude extremity (Downing, Judd, & Brauer, 1992), and attitude polarization (Brauer, Judd, & Gliner, 1995). Attitude accessibility was defined as the latency of response to an attitudinal inquiry. Thus, in certain contexts at least, researchers explicitly recognize that psychological measurement can have effects of its own.

In a more recent and related area of investigation, a number of independent research teams have also found important behavioral effects of "mere measurement" (e.g., Godin, Sheeran, Conner, & Germain, 2008; Greenwald, Carnot, Beach, & Young, 1987; Morwitz, Johnson, & Schmittlein, 1993; Sandberg & Conner, 2009; Sherman, 1980). For example, Sherman (1980) found that asking people to predict their future behavior yielded behavior that was consistent with those predictions and different from that of participants who were not asked to make predictions. Greenwald et al. (1987) found that asking people whether they intend to vote led to an increased likelihood of voting, which was mediated by behavioral intentions to vote. Morwitz et al. (1993), who coined the term *mere-measurement effect*, found that measuring an individual's purchase intentions changed subsequent purchase behavior (see also Morwitz & Fitzsimons, 2004). Similarly, Sandberg, & Conner (2009) found that measurement of anticipated regret led to higher cervical screening attendance rates.

In lieu of these findings, it makes sense that other researchers have raised concerns about measurement-induced disturbances in at least two domains of inquiry. Speaking to mediation testing, Lindsay & Anderson (2000) rightly pointed out that measurements that interfere with subsequent measurements violate the assumption of mediation testing. Also relevant are diary studies, which typically involve self-report instruments administered repeatedly for the examination of ongoing experience (Bolger, Davis, & Rafaeli, 2003; Iida, Shrout, Laurenceau, & Bolger, 2012; Moskowitz, Russell, Sadikaj, & Sutton, 2009; Reis & Gable, 2000; Wheeler & Reis, 1991). Recently, Iida et al. (2012) expressed concern about the degree to which repeated measurements in diary studies can change participants' experience and behavior.

Why focus on self-reported emotions?

Although our general concern applies to measurement of a broad range of psychological phenomena, we believe that emotion is a domain in which such concerns are especially important. Self-reports are a critically important mainstay of the measurement toolbox used by psychologists when studying emotions and their causes and consequences. Specifically, emotion researchers continue to examine the relevance of emotions for cognition, judgment, and behavior (e.g., Baumann & DeSteno, 2010; Rusting, 1999; Van Boven, Kane, McGraw, & Dale, 2010; White & Van Boven, 2012; see Lench, Flores, & Bench, 2011, for review). Therefore, potential biases in self-reported emotion intensity are relevant to a wide range of psychological research.

Moreover, the extant literature points to unanswered questions regarding the effects of self-report measurements of emotions. For instance, Kassam & Mendes (2013) experimentally manipulated self-reports of anger and shame induced in the laboratory. Half of the participants provided self-reports of their emotional states, whereas the other half completed a control questionnaire. Effects of self-reported emotions manifested in subsequent physiological measures, albeit selectively. Angry participants who reported their anger showed decreased cardiac output and increased total peripheral resistance, relative to those who did not. However, cardiovascular responses

amidst shame were not affected by self-reports. Unfortunately, the researchers did not directly examine effects on the measured construct itself.

Another example is research on affective labeling, especially studies in which affect labeling has been used as an emotion-regulation technique (Lieberman, Inagaki, Tabibnia, & Crockett, 2011; Ortner, 2015; Tabibnia, Lieberman, & Craske, 2008). In such studies, participants view a series of emotion-provoking stimuli and choose a word to identify the feeling. In particular, Lieberman et al. (2011) contrasted labeling of a series of emotional pictures in the labeling condition with merely looking at the same set of pictures in the control condition. However, it is possible that labeling hindered the development of an emotional response instead of regulating a full-blown response. Given that measurements of emotions are usually not conducted in this online fashion, we wished to more directly assess the effect of measuring emotions than is offered by studies on labeling.

Other researchers seem to at least implicitly acknowledge that measurements might have undesirable effects on research participants' subsequent emotions (e.g., Dunn & Schweitzer, 2005; Hunsinger, Isbell, & Clore, 2012; Keltner, Locke, & Audrain, 1993). Dunn & Schweitzer (2005) examined the influence of incidental emotions on trust in unrelated settings. In Study 1, the researchers induced different emotions but deliberately omitted conducting manipulation checks on their primary sample due to concerns that a manipulation check "would reduce the effects of our emotion induction" (p. 739). They instead tested their manipulation on a pilot sample. Thus, concerns about undesirable influences of measurements seem to have influenced some researchers' methodological decisions. Yet we are unaware of any direct investigations of the effects of self-report measurements on the psychological phenomena they are used to measure.

Overview

The prevalence of self-report emotion measures—plus evidence that some researchers have been concerned about measurements' possible unintended effects on emotions—suggests that emotion is a natural domain in which to look for evidence of observer effects in psychological measurement. We were directly interested in the effect of measurement on the measured quantity itself. Specifically, we examined the effect of measurement of emotion intensity on emotion intensity. A direct advantage is that our data allowed us to test the direction of the effect, that is, whether measurements reduce emotion intensity, maintain it, or increase it. In so doing, we expanded the range of emotions to include negative and positive emotions to test generalizability and/or detect boundary conditions for observer effects in self-reported emotion. The purpose of the current investigation, however, was to establish a reliable effect rather than to rule in a single mechanism.

Three experiments examined emotions of different valence (positive/negative) and nature (basic/complex) while utilizing different emotion induction paradigms. Experiments 1 and 2 examined negative emotions—sadness and envy, respectively. Experiment 3 examined a positive emotion (happiness). The basic emotions of sadness and happiness were induced using affective story recall. The complex emotion of envy was induced via comparisons with a high- versus low-status peer. Despite these differences, the methodology was similar across all experiments. Each included an emotion induction task followed by a final self-report measurement of emotions. We manipulated the number of measurements before the final measurement. Unrelated reading tasks were included as fillers. In Experiment 3, we complemented final self-reports of emotion intensity with measurements of behavior relevant to happiness.

Hypotheses

At the core of our investigation was the idea that self-report measurements can influence ongoing emotional experience. Our first prediction was based on the finding that simply paying attention to emotion can lead to a decrease in negative affect (Thompson et al., 2011). We hypothesized that more measurements would weaken negative emotions. We call this the "numbing effect" and examined it in all experiments. The hypothesized numbing effect is also consistent with the idea of habituation in diary studies (Bolger et al., 2003).

In Experiment 3, we tested whether the patterns observed with negative emotions would hold with a positive emotion, namely, happiness. We considered this study to be more exploratory due to the unclear theoretical and empirical support regarding positive affect. For example, although Thompson et al. (2011) found that emotion attention reduced negative affect, they did not find similar effects on positive affect. In addition, opposing forces might be at play in the case of positive emotions. For instance, people are prone to maintaining positive emotions (Tamir, 2009; Volokhov & Demaree, 2010). Although we might

expect that measurements would reduce the intensity of positive emotions, this effect might be less pronounced than the effect on negative emotions because the motivation to maintain the positive emotions might counter the effect of measurement. Finally, we examined the happiness-specific task of impression formation, with the intent of investigating the effects of prior self-report measures on measures other than self-reports of emotion.

Experiment 1: Sadness

Method

Participants and design

This study comprised 198 participants (117 female) from a large midwestern university who earned research credit toward a course requirement as compensation for their participation. Participants were randomly assigned to either a neutral mood induction condition or one of the three mood (sadness) induction conditions. The three mood induction conditions differed in the number of self-report measurements before the final measurement (we use self-explanatory labels to refer to these conditions, namely, six-measurement, two-measurement, and one-measurement). Emotion intensity was measured over a 25-min period after initial emotion induction.

Comparison of the six- and the one-measurement conditions allowed us to examine the effect of measurements on final emotion intensity. Because we expected no differences in the initial effects of our mood induction, we hypothesized that at Time 1 the six- and the two-measurement conditions would show comparable levels of sadness. Moreover, a comparison between the six- and the two-measurement conditions at Time 6 allowed us to examine the effect of measurements on change in emotion intensity between initial and final assessments.

Materials and procedures

Participants completed the study online. First, participants recalled and described an experience from their recent past. In the neutral group, this experience was a trip to the grocery store. In the three emotion groups, this experience was a recent episode of sadness. Participants typed descriptions of their experiences. Next, participants answered some objective questions about their experience, for instance, "How many months and years have passed since the occurrence of the experience you just described?" Between emotion induction and the final measurement, participants completed five unrelated filler tasks (e.g., they read and answered questions about articles from websites of National Geographic and the Discovery Channel). Filler tasks were interspersed with measurements of emotional intensity contingent upon the condition, as described next.

In the neutral-mood and the six-measurement conditions, emotion intensity was measured five times before the final measurement. The first measurement occurred immediately after emotion induction. Consecutive measurements were separated by 5-min intervals. In the two-measurement condition, emotions were measured once right after emotion-induction and then 25 min later. In the one-measurement condition, emotions were measured only at the end of the 25-min interval after induction. To avoid confounding with time, an additional 30 s were allocated for each filler task in the case of participants who did not receive a measurement. Thus, the conditions differed in the number of emotion measurements after emotion induction. We denote measurements by the time at which they occurred (e.g., Time 6 measurement).

Each measurement of sadness employed the Positive and Negative Affect Schedule–Expanded Form (Watson & Clark, 1994). On a scale of 1 (*not at all*) to 5 (*strongly*), participants rated the extent to which they felt *sad, blue, downhearted, alone,* and *lonely*. Responses were aggregated to compute a sadness score. Participants also completed the Alertness, Shyness, and Fatigue subscales to minimize transparency and demand characteristics, resulting in 17 items.

Results

Means and standard deviations of sadness have been reported in Table 1. First, we examined the manipulation of sadness via the affective recall task. Sadness scores at Time 1 ($\alpha = 0.92$) were submitted to a one-way analysis of variance (ANOVA; $\eta^2 = 0.07$). Planned contrasts revealed that Time 1 sadness in the six- and the two-measurement conditions taken together ($M = 2.30$, $SD = 1.09$) was stronger than in the neutral mood condition ($M = 1.70$, $SD = .89$; $d = 0.60$). Time 1 sadness in the six-measurement condition ($M = 2.20$, $SD = 0.93$) was fairly similar to the two-measurement condition ($M = 2.39$, $SD = 1.18$; $d = -0.18$). Second, we examined the effect of measurement on sadness at Time 6. Sadness scores at Time 6 ($\alpha = 0.92$) were submitted to a one-way ANOVA. Planned contrasts revealed that Time 6 sadness in the six-measurement condition ($M = 1.48$, $SD = 0.72$) was weaker than in the two- and the one-measurement

Table 1. Mean emotion intensity of induced emotion.

Experiment	Emotion induction	No. of measurements	Time 1	Time 2	Time 3	Time 4	Time 5	Time 6
1	Sadness	6	2.20 (0.93)	1.68 (0.72)	1.63 (0.79)	1.49 (0.64)	1.55 (0.80)	1.48 (0.74)
	Sadness	2	2.39 (1.18)	—	—	—	—	1.94 (1.14)
	Sadness	1	—	—	—	—	—	1.82 (0.83)
	Neutral	6	1.70 (0.89)	1.58 (0.82)	1.51 (0.71)	1.48 (0.67)	1.45 (0.66)	1.42 (0.62)
2	Envy	6	4.78 (2.54)	3.68 (2.66)	3.50 (2.88)	3.39 (2.97)	3.09 (2.80)	3.03 (2.72)
	Envy	1	—	—	—	—	—	4.62 (2.80)
	Neutral	6	1.40 (0.80)	1.48 (1.07)	1.34 (0.96)	1.30 (0.82)	1.36 (0.97)	1.28 (0.72)
3	Happiness	4	7.00 (2.26)	5.92 (1.96)	5.55 (2.21)	4.94 (1.93)	—	—
	Happiness	1	—	—	—	4.83 (1.61)	—	—

Note. Standard deviations appear in parentheses.

conditions taken together (M = 1.85, SD = 0.98; d = 0.43, η^2 = 0.04). Thus, as hypothesized, sadness was weaker after the same time delay postinduction if it had been measured multiple times rather than only once or twice.

In addition to reported intensity, we also examined the decline in sadness by computing the difference in sadness between Time 1 and Time 6. Note that this was possible only for conditions in which sadness was measured prior to Time 6. Overall, sadness in the six-measurement condition decreased from Time 1 (M = 2.20, SD = 0.93) to Time 6 (M = 1.48, SD = 0.72; d = 0.86). The decrease in the six-measurement condition (M = 0.76, SD = 0.89) was larger relative to the two-measurement condition (M = 0.40, SD = 0.64; d = 0.46). Thus, over the same period, sadness declined at a faster rate when more measurements were administered. Next, we examined the decrease in sadness between Time 1 and Time 2. In the six-measurement condition, the decrease between Time 1 and Time 2 (M = 0.52, SD = 0.71) was only moderately smaller than the overall decrease between Time 1 and Time 6 (M = 0.76, SD = 0.89; d = 0.29) and not substantially different from the overall decrease in the two-measurement condition between Time 1 and Time 6 (M = 0.40, SD = 0.64; d = 0.18). This suggests that one prior measurement might be sufficient to bring about a noticeable effect of measurement on emotions.

Experiment 2: Envy

Although the findings of Experiment 1 supported our hypothesis regarding the effect of measurements on emotion intensity, we wanted to rule out some alternate explanations. Given high face-validity of Experiment 1, it is possible that participants tried to make sense of the frequent measurements in the context of the induced emotion of sadness. For instance, participants might have reported weaker emotions in subsequent measurements in attempts to be consistent with lay understanding about emotions (e.g., emotions become weaker over time). Therefore, in Experiment 2, we used a cover story to draw attention away from the purpose of the study and thus minimize the potential role of demand characteristics.

Furthermore, memory biases and the personal relevance of the recalled sadness episodes in Experiment 1 could have driven the effect of measurements on subsequent emotion. We address these concerns in the Experiment 2 via a different approach to emotion induction. We induced emotion directly in the lab instead of relying on affective recall. Finally, Experiment 2 focused on a different, more complex negative emotion (envy) to broaden the generalizability of our investigation.

Method

Participants and design

This study comprised 137 participants (49 female) from a large midwestern university who earned research credit toward a course requirement as compensation for their participation. Participants were randomly assigned to either a neutral-emotion condition or one of two emotion (envy) conditions. The emotion conditions differed in the number of measurements administered prior to the final measurement of emotional intensity. We use self-explanatory labels to refer to these conditions, namely, six-measurement and one-measurement. Emotion intensity was measured over a 25-min period after initial emotion induction.

Materials and procedures

Upon participants' arrival at the laboratory, the experimenter introduced the study as examining "how perceptions and feelings toward the content of news stories may change based on how the information is presented." Furthermore, participants were told that one interest was to simulate the process of taking breaks while receiving information online. Participants were told that although people in the study would receive information across many different media, they would read an article from the university newspaper followed by articles from "international websites"

(which served as filler tasks). Participants were seated at a computer on which the first article appeared. The online version of the article was developed in collaboration with the university newspaper, thus rendering it indistinguishable from any "real" article that the students would normally encounter.

This first article was ostensibly from a series of interviews profiling individual students on campus. The fictional individual in the article was of the same gender as the participant. In the neutral-emotion condition, the content of the article left an impression that the individual was average (an average-looking, mediocre student with some financial struggles and an unexceptional social life). In the two envy conditions (six- and one-measurement), the individual was superior (a physically attractive, excellent student with upper middle-class parents and an exciting social life). Between emotion induction and the final measurement, participants completed five unrelated filler tasks as in Experiment 1. Filler tasks were interspersed with measurements of emotional intensity contingent upon the condition. In the neutral-mood and six-measurement conditions, emotion intensity was measured five times before the final measurement. The first measurement occurred immediately after induction. Consecutive measurements were separated by 5-min intervals. In the one-measurement condition, emotions were only measured at the end of the 25-min interval after induction. To avoid confounding with time, additional 30 s were allocated for each filler task in the case of participants who did not receive a measurement. Comparison of the final emotion intensity in the six- and the one-measurement conditions allowed us to examine the effect of measurements on emotions.

Each measurement of emotions was administered covertly. Participants were instructed that "comprehension may be influenced by feelings" and that it was important to measure how they felt toward the individual in the article. Following these instructions, participants were asked to rate their feelings on 32 emotional reactions (see Krizan & Johar, 2012). This presentation was designed to decrease the likelihood that participants would guess that assessing envy (or any other specific emotion) was central to the study goals. We created an Envy index based on the responses to *envious of*, *jealous of*, and *resentful envy toward*. A scale of 0 (*not at all*) to 10 (*a great amount*) was used for responses.

Results

Means and standard deviations of envy have been reported in Table 1. First, we examined the manipulation of envy via the newspaper article task. Time 1 envy (α = 0.87) in the neutral condition (M = 1.40, SD = 0.80) was weaker than Time 1 envy in the six-measurement condition (M = 4.78, SD = 2.54), confirming that we successfully manipulated envy (d = 1.79). Second, we examined the effect of measurement on envy at Time 6. Time 6 envy scores (α = 0.90) were submitted to a one-way ANOVA (η^2 = 0.07). Time 6 envy in the six-measurement condition (M = 3.03, SD = 2.72) was weaker than Time 6 envy in the one-measurement condition (M = 4.62, SD = 2.80; d = 0.57). Thus, consistent with our hypothesis and the results of Experiment 1, envy was lower due to more measurements of the emotion. Figure 2 presents the emotion-measurements across time in different groups. Note that Time 1 envy in the six-measurement condition is comparable to the Time 6 envy in the one-measurement condition.

Next, we more closely examined the decrease in envy. First, there was a sizeable overall decrease in envy in the six-measurement condition between Time 1 (M = 4.78, SD = 2.54) and Time 6 (M = 3.03, SD = 2.72; d = 0.66). Second, in the six-measurement condition, this decrease between Time 1 and Time 6 (M = 1.93, SD = 2.08) was moderately greater than the decrease between Time 1 and Time 2 (M = .96, SD = 1.86; d = 0.49). Therefore, later measurements after the first one continued to decrease the intensity of envy. Third, Time 2 envy in the six-measurement condition (M = 3.68, SD = 2.66) was smaller than Time 6 envy in the one-measurement condition (M = 4.62, SD = 2.80; d = 0.34). Consistent with the findings of Experiment 1, it seems that one prior measurement might be sufficient to bring about a noticeable effect of measurement on emotions.

Discussion

An intriguing finding was that Time 6 envy in the one-measurement condition was comparable to Time 1 envy in the six-measurement condition. It seems to suggest that the absence of measurements may have served to "preserve" emotional intensity at its initial level. However, there could be another possibility. The purpose of the cover story was to divert focus away from the purpose of the study. In doing so, the cover story could have prevented a thorough evaluation of the envy-provoking stimuli (in terms of the high-status social comparison target). As such, participants might not have fully experienced envy until the first measurement forced them to think about it. Hence, when participants in the no-measurement condition

reported their emotions for the first time, at Time 6, their psychological experience (not just emotion intensity as measured) was comparable to Time 1 envy in the five-measurement condition. Regardless of which of these competing explanations is true, these findings nevertheless support our hypothesis that measurements of emotions can have a systematic effect on the ongoing emotional experience.

Experiment 3: Happiness

Experiment 3 extended the findings of the first two experiments in multiple ways. First, we examined the positive emotion of happiness in addition to negative emotions. We hoped to test generalizability of the findings so far. As discussed earlier, we expected that negative emotions would be more strongly affected by repeated measurements owing to the motivation to maintain positive emotions (Tamir, 2009). Second, similar to Hunsinger et al. (2012), we measured individual (dispositional) differences in attention to emotion to more closely examine the effect of repeated measurements.

We assume that repeated measurements draw attention to emotion and that this enhanced attentional focus would carry over to subsequent tasks, as shown by prior research (e.g., Forster, Friedman, & Liberman, 2004; Trope & Liberman, 2003; Wakslak, Trope, Liberman, & Alony, 2006; see Forster & Dannenberg, 2010). Furthermore, happiness has been found to promote the use of accessible cognitive responses (Clore & Huntsinger, 2007). Therefore, we expected to observe the effects of enhanced attention in downstream processes. However, the increase in attentional focus (owing to measurement) should be stronger in people who score low on baseline attention to emotion. Thus, we hypothesized that individual differences in attention to emotion should moderate the effect of measurement.

We were also concerned that altered reports of emotion intensity may or may not necessarily denote an actual change in the underlying emotional experience. To more closely examine the underlying emotional experience, we complemented the self-report measures with an indirect measure of emotion in Experiment 3. However, we do not wish to imply that indirect measures necessarily offer greater validity while measuring cognitive processes (Gawronski & Hahn, in press).

For the indirect measure, we chose to examine mood effects on stereotyping because of the preponderance of evidence for such effects (Bless, 2000; Bodenhausen, Kramer, & Susser, 1994; Isbell, 2004). For example, Hunsinger et al. (2012) had participants complete an impression formation task that included both categorical (e.g., trait labels) and behavioral information about a target individual. When participants were led to focus on their immediate psychological state (i.e., current affective experience), happy moods led to judgments reflecting behavior-level information. We expected similar results in our Experiment 3.

We maintain that repeated measurements tend to draw attention toward the emotional experience of an individual. In an impression formation task, repeated measurements should nudge individuals away from relying on categorical information and more toward behavioral-level information. We hypothesized that repeated measurement of emotion would thus reduce stereotype-consistent impression formation. Again, we expected that individual (dispositional) differences in attention to emotion would moderate the effect of repeated measurements.

Method

Participants and design

The study comprised 215 participants (116 female) from a medium-size midwestern university who earned research credit toward a course requirement as compensation for their participation. The affective story recall paradigm used in Experiment 1 was used to induce happiness in all participants. Note that we measured sadness in this experiment but did not induce it. In the first two experiments, emotion intensity stabilized by the fourth assessment. Therefore, to streamline our method, participants in Experiment 3 were randomly assigned to either a one-measurement or a four-measurement condition. Emotion intensity in both conditions was thus measured over a corresponding 15-min period after initial emotion induction. Comparison of groups allowed us to examine the effect of measurements on final emotion intensity and on the impression formation task.

Materials and procedures

The order in which emotion induction, measurements, and filler tasks appeared was similar to that in Experiment 1. In the four-measurement condition, emotion intensity was measured at approximately 0, 5, and 10 min after emotion induction, and again at the conclusion of 15 min. In the one-measurement condition, emotions were measured only at the conclusion of 15 min. To avoid confounding with time, an

additional 30 s were allocated for each filler task in the case of participants who did not receive a measurement. Prior to the final measurement, all participants completed an impression formation task. It was critical to administer the impression formation task before the final measurement. We anticipated that the final measurement would increase attention to emotions even in the one-measurement group, thereby making it harder to observe the effect of attention in the impression formation task.

Happiness measurements employed the Joviality subscale of the Positive and Negative Affect Schedule–Expanded Form (Watson & Clark, 1994). On a scale of 0 (*not at all*) to 10 (*a great amount*), participants rated the extent to which they felt *happy, joyful, delighted, cheerful, excited, enthusiastic, lively,* and *energetic*. Responses were aggregated to compute a happiness score. Participants also completed the Sadness and Alertness subscales described in Experiment 1.

Finally, participants completed the emotion-attention measure adapted by Hunsinger et al. (2012). This measure includes items from the Emotional Creativity Scale (Averill, 1999) and the Meta-Mood Scale (Salovey, Mayer, Goldman, Turvey, & Palfai, 1995). Three items each were taken from Emotional Creativity Scale (e.g., "I think about and try to understand my emotional reactions") and the Meta-Mood Scale (e.g., "I often think about my feelings"). We used an 11-point scale from 0 (*none at all*) to 10 (*a great amount*; $\alpha = 0.84$).

Impression Formation Task. Participants read a story about an individual described as highly extroverted (Hunsinger et al., 2012). The story presented a list of behaviors, only half of which were consistent with the category label (i.e., stereotype of an extroverted individual). After reading the story, participants rated the extent to which different stereotype-relevant terms described the target individual on a scale from 0 (*none at all*) to 10 (*a great amount*). Impressions of extroversion were computed as the aggregate of responses to *withdrawn, shy, a loner, quiet, talkative, self-confident, sociable,* and *outgoing*. The first four items were reverse-coded ($\alpha = 0.76$). Participants also reported how carefully they read the story, on the same 0–10 scale.

Results

Carefulness. Carefulness scores were fairly similar in the repeated measurements condition ($M = 7.70$, $SD = 1.40$) and the single measurement condition ($M = 7.50$, $SD = 1.30$), $d = 0.15$.

Self-reported emotion. Sadness at Time 4 ($\alpha = 0.89$) in the four-measurement group ($M = 2.59$, $SD = 1.64$) was lower than sadness in the one-measurement group ($M = 3.15$, $SD = 1.70$; $d = 0.34$). To examine the effects of our manipulation and dispositional emotion-attention, ratings of sadness at Time 4 were submitted to a hierarchical regression with the condition, emotion-attention, and their product as predictors. We observed a main effect of condition ($\beta = -0.35$, $\eta^2 = 0.03$) and an interaction effect ($\beta = 0.18$, $\eta^2 = 0.01$) but no main effect of emotion-attention ($\eta^2 < 0.01$). Those low on emotion-attention (-1 *SD*) showed much lower sadness scores as a result of repeated measurement (simple slope test: $\beta = -0.53$). Those high on this dimension were not affected by repeated measurement as strongly (simple slope test: $\beta = -.17$). Thus, relative to those who scored high on emotion-attention, those who scored low on emotion-attention showed a stronger reduction in sadness, owing to repeated measurements. Overall, sadness in the four-measurement condition declined from Time 1 ($M = 3.71$, $SD = 2.36$) to Time 6 ($M = 2.59$, $SD = 1.64$; $d = 0.40$).

In contrast to findings involving sadness, happiness at Time 4 ($\alpha = 0.96$) in the four-measurement group ($M = 4.94$, $SD = 1.93$) was almost identical to happiness in the one-measurement group ($M = 4.83$, $SD = 1.64$; $d = 0.06$; see Figure 3). To examine the effect of our manipulation and dispositional emotion-attention, ratings of happiness at Time 4 were submitted to hierarchical regression with the condition, emotion-attention, and their product as predictors. There were no main effects of measurement or emotion-attention ($\eta^2 < 0.01$). However, an interaction was observed ($\beta = 0.24$, $\eta^2 = 0.02$). Those low on emotion-attention (-1 *SD*) showed lower happiness scores as a result of repeated measurement (simple slope test: $\beta = -0.19$). However, those high on this dimension showed higher happiness scores as a result of measurement (simple slope test: $\beta = 0.31$). Thus, data suggest, albeit weakly, that the manipulation of measurements lowered happiness for those who scored low on emotion-attention, relative to those scoring high on emotion-attention.

Impression-formation task. To examine the effect of our manipulation and dispositional emotion-attention on the likelihood of stereotype-consistent impression formation, impressions of extraversion were submitted to hierarchical regression with the condition, emotion-attention, and their product as predictors. The carefulness score was used as a covariate. There were no main effects of measurement or emotion-attention ($\eta^2 < 0.01$). As expected, emotion-

attention interacted with the measurement manipulation to shape impressions of extraversion (β = 0.32, η^2 = 0.03). Specifically, emotion-attention moderated the effect of the measurement manipulation on impression-formation (Aiken & West, 1991). Those low on emotion-attention (−1 *SD*) showed weaker impressions of extroversion as a result of repeated measurement (simple slope test: β = −0.47). However, those high on this dimension were not particularly affected (simple slope test: β = 0.17). As mentioned earlier, the carefulness item was fairly similar across the two levels of condition. Moreover, the interaction effect was still observed even when carefulness was not included as a covariate (β = 0.24, η^2 = 0.02).

Next, we more closely examined the decrease in happiness. First, happiness in the four-measurement condition showed a sizeable decline from Time 1 (*M* = 7.00, *SD* = 2.26) to Time 6 (*M* = 4.94, *SD* = 1.93; *d* = 0.98). Second, in the four-measurement condition, happiness at Time 1 (*M* = 7.00, *SD* = 2.26) was moderately greater than Time 2 (*M* = 5.92, *SD* = 1.96; *d* = 0.51). Therefore, roughly half of the decline in happiness was observed by Time 2. Consistent with the findings of Experiments 1 and 2, it seems that one prior measurement might be sufficient to bring about a noticeable effect of measurement on emotions.

Discussion

Participants reported both sadness and happiness. Although Experiment 3 strongly replicated the findings of Experiments 1 and 2 with regard to sadness, repeated measurement was not found to influence happiness overall. This suggests an interesting and important potential boundary condition (see also Thompson et al., 2011). Notably, we did observe weak evidence for the moderating role of attention to emotions. Participants scoring low on attention to emotion were more likely to demonstrate reduced happiness (and sadness) owing to repeated measurement.

Furthermore, the impression formation task helped to advance our understanding of how measurements affect ongoing emotional experience. Although we did not observe a main effect of our manipulation on impression-formation, the manipulation interacted with attention to emotion. The degree to which people naturally adopt behavioral-level processing is partially dependent upon their attention to their affect-eliciting experiences (Hunsinger et al., 2012). Assuming that self-report measurement of emotion increases attention to emotion, it should also lead to greater behavioral-level processing. However, people who exhibit already-high trait scores on attention to emotion should thus be relatively unaffected on this dimension. It thus makes sense that participants in Experiment 3 who reported high trait scores of attention to emotion did not demonstrate higher behavior-level (or lower categorical-level) processing as a result of more measurements. In contrast, participants with low trait attention to emotion showed greater behavioral-level processing (and lesser category-level processing) as a result of more measurements. We thus infer that measurements interfere with ongoing emotional experience by priming more attentive and reflective cognitive styles, but only among individuals who are not already highly attentive to their emotional experience.

A closer examination of the rate of decay of different emotions suggests an interesting possibility. We calculated effect sizes for the change in emotion intensity between the first and the final measurements across Experiments 1, 2, and 3. We found moderate to strong effect sizes: Cohen's *d*s were 0.86 (sadness in Experiment 1), 0.66 (envy in Experiment 2), 0.98 (happiness in Experiment 3), and 0.4 (sadness in Experiment 3). Given that happiness declined at a faster rate, it is plausible that it was too late to detect the effect of measurement by Time 4.

General discussion

Psychology's status as a scientific field is sometimes criticized because psychological phenomena are difficult if not impossible to perfectly measure. Although psychological researchers themselves understand that measurements are fallible (Cole & Preacher, 2014; Schmidt & Hunter, 1996), there is a dearth of direct evidence regarding measurement-induced perturbations. Thus, the field has a surprisingly limited understanding of the ways in which measurements can affect the phenomenon being studied (but see Knowles et al., 1996; Sharpe & Gilbert, 1998; Shrout et al., 2018). Furthermore, what prior evidence exists seems to have had little cautionary effect, because, as mentioned previously, self-reported emotions remain widely used (and sometimes with multiple measurements) with little apparent concern about potential self-contamination. Although we hope that it is not the case, it is nonetheless possible that our findings threaten the validity of a large swath of research using such methods.

With such concerns in mind, the objective of the current research was to better understand ways in

which psychological measures limit empirical conclusions. We focused on self-reports partly because of their widespread use and importance to psychological research, despite their well-documented limitations (e.g., Donaldson & Grant-Vallone, 2002; Nisbett & Wilson, 1977; Podsakoff, MacKenzie, Lee, & Podsakoff, 2003; Spector, 1994), but also because of their inherent tendency to increase individuals' attention to their own psychological experience.

Although psychologists acknowledge that self-reports can only help uncover processes that are accessible to conscious awareness (Nisbett & Wilson, 1977), criticism of self-report measures typically stems from the view that biases in self-reports constitute additive method variance (Podsakoff et al., 2003). This might lead one to believe that self-reports merely hinder accurate measurement, but our findings suggest that self-reports interfere on an even deeper level: Such measures can actively alter the psychological phenomena being observed. For instance, across all three experiments, self-report measures of emotion reduced the intensity of negative emotions.

Why did self-reports of emotion intensity change emotion intensity?

As mentioned earlier, our primary goal was to examine the self-contaminating effects of measurements given the lack of research that directly does so. Across three studies, we found that repeated measurements tend to weaken the intensity of negative emotions but not positive emotions. Experiment 3 hinted at the role of moderator of this effect. However, adequately uncovering the underlying mechanisms was beyond the scope of this investigation. In the absence of additional data, we offer speculations about potential underlying pathways that may have driven the self-contaminating nature of measurements.

First, simply paying attention to emotions may systematically alter them, contingent upon valence. Thompson et al. (2011) showed that attention to emotion predicted a decrease in negative affect but did not change positive affect. Given that attention may open the door to regulation of emotion, the motivation to regulate or maintain certain emotions might also play a role. In general, people are prone to regulating negative emotions while maintaining positive emotions (Volokhov & Demaree, 2010). This might explain why we found stronger reduction in negative (vs. positive) emotions due to repeated measurements.

Second, self-report measurements of emotion could facilitate repetitive thinking and influence the emotional state, contingent upon how people think about their emotions (Watkins, 2008). Reappraisal might serve to regulate the emotion, thus accelerating the decline of emotional intensity (Kalokerinos, Greenaway, & Denson, 2015; Ray, McRae, Ochsner, & Gross, 2010; Urry, 2009; Webb, Miles, & Sheeran, 2012). On the other hand, research on rumination suggests that repetitive thinking stemming from measurement could serve to maintain the emotion (Nolen-Hoeksema, Morrow, & Fredrickson, 1993; Ray, Wilhelm, & Gross, 2008; Trapnell & Campbell, 1999; Whitmer & Gotlib, 2013). There are, therefore, competing arguments for how repetitive thinking from self-reported affect might influence emotion intensity. Regardless, the present data did not show maintenance of emotion by virtue of rumination.

Third, appraisal theories suggest that novelty of emotional stimuli is important for a strong emotional response. Self-report measurements may have reduced novelty via habituation owing to increased exposure to emotional stimuli. Consequently, emotional stimuli might elicit only weaker emotional reactions (Scherer, Schorr, & Johnstone, 2001). Fourth, it is also possible that the observed effects emanated from demand characteristics as suggested by the findings of Shrout et al. (2018). When participants made their first rating, they likely compared their current feeling to the time before the emotion was induced. Thus, they reported relatively strong emotions. By the second measurement, part of that evaluation may have been relative to the previous measurement. They may have provided a weaker rating because they did not feel worse than the previous measurement. In a similar vein, it is also plausible that participants became less invested in the study with each additional measurement. This might also explain the reduction of emotion intensity owing to measurement. However, this reasoning is not supported by Study 3, where the effect of the manipulation was not identical across PA and NA.

Implications

We suggest that repeated measurement of emotional intensity is a worthy area of future study as an emotion-regulation strategy. Measurement of negative emotions may facilitate regulation in a manner similar to labeling affective experiences and writing about emotions. Compared to these emotion-regulation interventions, self-report of emotion intensity can be utilized as a less obvious (i.e., face-valid) tool that is less likely to elicit psychological reactance (Brehm, 1966) and is quick to administer. Writing about emotions is, comparatively, a more resource-intense

strategy. Given the simplicity of repeated measurements of emotion, they could be a component worth adding to interventions that do not already incorporate repeated measurements.

While adding to the literature on methodological concerns about reliance upon self-reports, our findings pose a word of caution for researchers. By no means do we condemn the use of self-reports in the emotion literature. However, we urge fellow researchers to exercise caution and be mindful of the potential effects that measurement can have on the underlying construct in at least three contexts. First, researchers relying on mediation testing should take note. Across all experiments, we found a sizeable change in emotions after the first measurement. Thus, the measurement of a potential mediator, even once, could render the mediator weaker or stronger owing to measurement. This has serious implications for observed relations between the independent variable, the mediator, and the dependent variable. Thus, our findings add to the available criticisms of mediation testing (Grice, Cohn, Ramsey, & Chaney, 2015; Kline, 2015; Lindsay & Anderson, 2000; Tate, 2015; Thoemmes, 2015).

Second, our findings suggest why the use of manipulation checks might be a bad idea. Similar to mediation testing, measurement intended to perform a manipulation check could perturb the underlying psychological experience, thereby compromising the effect on the main dependent variable. Consistent with previous research, our findings recommend a careful analysis of the benefits vs. costs of using manipulation checks (Kidd, 1976; Sawyer, Lynch, & Brinberg, 1995; Trafimow & Rice, 2009).

Third, researchers interested in the time course of emotions and those who rely on diary studies and survival analysis should also take note. However, such studies often involve longer intervals of time between consecutive measurements (longer than 5 min). As such, we are able to offer only broad suggestions without more research on the generalizability of our findings to diary studies.

The findings of Experiment 3 also point toward a potential moderator of the effect that emotions have on stereotyping. Previous research has found that incidental emotions can be important for stereotyping of certain groups (Bodenhausen, Mussweiler, Gabriel, & Moreno, 2001; Dasgupta, DeSteno, Williams, & Hunsinger, 2009; Huntsinger, Sinclair, Dunn, & Clore, 2010). The impression formation task used in Experiment 3 revealed that attention to emotion shapes the degree to which impressions are stereotype consistent or inconsistent. Thus, our findings add to the literature at the interface of incidental emotions and stereotypes or prejudice (Hunsinger et al., 2012; Huntsinger et al., 2010).

Our research also has implications for real-world contexts in which people are often repeatedly asked to report their feelings, such as in areas of patient or customer satisfaction and loyalty. For example, research has shown that when they take the opportunity to complain, dissatisfied customers may become more satisfied (Nyer, 2000) and/or exhibit an increase in loyalty (Umashankar, Ward, & Dahl, 2017). Although there are surely multiple factors contributing to these effects, our data suggest that asking individuals who have negative feelings about a patient or customer experience to report their emotions may reduce those emotional aspects of dissatisfaction. Future research could therefore investigate behavioral consequences of repeated measurement in consumer contexts, such as whether people are more likely to return to a business or care provider with which they have had a dissatisfying experience if they are first asked to report their negative emotions regarding the experience at least once, if not multiple times.

Future directions

Our investigation provides a first step toward understanding the effect of measurements on measured psychological phenomena such as emotions. Our hope is that this investigation spawns a broad collection of follow-up investigations to more completely understand the psychological experience of various types of measurements. It is possible, if not likely, that the self-contaminating effects observed in the present experiments are observable in other domains beyond emotion (such as motivation or attitudes) and, perhaps, to measurements beyond direct self-report. We also understand that many readers may desire that future investigations pursue direct evidence for the specific mechanism(s) that may underlie the present phenomenon. Although we do not discourage such investigations, we agree with Trafimow (2012) that the pursuit of explanatory mechanisms for their own sake may distract scientific scholars from more important goals such as the development of unifying theories. In our case, we are most interested in the phenomenon itself because of its cautionary value to researchers who use self-reports of measurements. To this end, the search for mechanisms may yield interesting and valuable new insights, but exploration into a broader range of domains and measurement may be

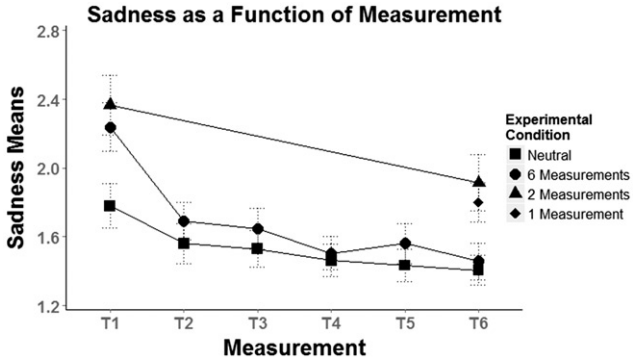

Figure 1. Intensity of sadness across experimental groups at different time points. *Note.* Dashed error bars denote 95% confidence intervals around the mean. Means in the same condition are connected by solid lines.

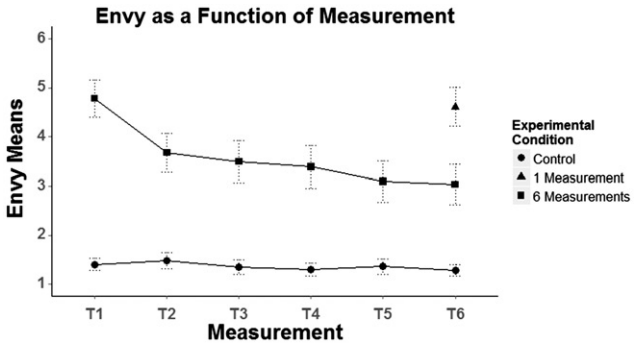

Figure 2. Intensity of envy across experimental groups at different time points. *Note.* Dashed error bars denote 95% confidence intervals around the mean. Means in the same condition are connected by solid lines.

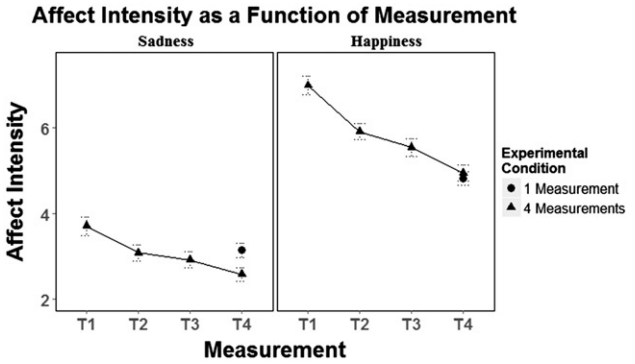

Figure 3. Intensity of sadness and happiness across experimental groups at different time points. *Note.* Dashed error bars denote 95% confidence intervals around the mean.

more important both theoretically (in terms of offering guidance toward unification) and methodologically (in terms of offering prescriptions for better measurement practices).

More specific to our experiments, we knowingly confounded the number of measurements (i.e., how many measurements took place total) with the frequency of measurements (i.e., the temporal proximity of measurements to induction and to one another) to keep the temporal distance between emotion induction and final measurement constant. However, as a result of this methodological decision, it remains unclear whether the observed effects of measurements on emotion resulted from the numerousness or the frequency of measurements. Although our current conceptualization of the process at play favors numerousness over frequency, it may be that both numerousness and frequency play a role. For example, the influence of multiple measurements might be diminished if these measurements are spread too far apart in time. Our data shed some light in this domain. As per Figures 1–3, the biggest decreases in emotion intensity were observed between Time 1 and Time 2.

Also, in our experiments, we adopted a restricted approach to emotions by focusing solely on emotional intensity, which is a simplistic indicator of emotional experience. In addition, duration and frequency are also characteristic of the ongoing experience of emotions (Verduyn & Brans, 2012). The field would benefit from understanding how measurements affect these other aspects of emotions. For instance, would negative emotions would run a shorter course owing to being measured, or would they occur less frequently?

Conclusion

There is broad consensus that psychological measurements are not perfect. Commonly understood limitations of measurements are biases in memory and self-awareness. As such, measurements are frequently viewed as imperfect but innocent tools for capturing a glimpse of psychological phenomena. We question this innocence, and in a series of experiments we examined this question empirically. In the context of self-reports of emotional intensity, we showed that measurements can actively interfere with the observed phenomenon, thus not only imperfectly capturing the psychological experience but potentially contaminating it indelibly. Our findings offer new theoretical insights and offer a cautionary tale to psychological scientists and others who use self-report measures. However, at the broadest level, we hope that these findings draw attention to the importance of treating measurements as not merely inert psychological instruments but as psychological experiences in themselves.

Disclosure statement

No potential conflict of interest was reported by the authors.

References

Adair, J. (1984). The Hawthorne Effect: A Reconsideration of the Methodological Artifact. *Journal of Applied Psychology*, 69, 334–345.

Aiken, L. S., & West, S. G. (1991). *Multiple Regression: Testing and interpreting interactions*. Newbury Park, CA: Sage.

Averill, J. R. (1999). Individual differences in emotional creativity: Structure and correlates. *Journal of Personality*, *67*(2), 331–371.

Baumann, J., & DeSteno, D. (2010). Emotion guided threat detection: Expecting guns where there are none. *Journal of Personality and Social Psychology*, *99*(4), 595–610.

Bless, H. (2000). The interplay of affect and cognition: The mediating role of general knowledge structures. In J. P. Forgas (Ed.), *Feeling and thinking: The role of affect in social cognition* (pp. 201–222). New York, NY: Cambridge University Press.

Bodenhausen, G. V., Kramer, G. P., & Susser, K. (1994). Happiness and stereotypic thinking in social judgment. *Journal of Personality and Social Psychology*, *66*(4), 621–632.

Bodenhausen, G. V., Mussweiler, T., Gabriel, S., & Moreno, K. N. (2001). Affective influences on stereotyping and intergroup relations. In J. P. Forgas (Ed.), *Handbook of affect and social cognition* (pp. 319–343). Mahwah, NJ: Erlbaum.

Bolger, N., Davis, A., & Rafaeli, E. (2003). Diary methods: Capturing life as it is lived. *Annual Review of Psychology*, *54*, 579–616.

Brauer, M., Judd, C. M., & Gliner, M. D. (1995). The effects of repeated attitude expressions on attitude polarization during group discussions. *Journal of Personality and Social Psychology*, 68, 1014–1029.

Brehm, J. W. (1966). *A theory of psychological reactance*. New York, NY: Academic Press.

Buks, E., Schuster, R., Heiblum, M., Mahalu, D., & Umansky, V. (1998). Dephasing in electron interference by a 'which-path' detector. *Nature*, *391*(6670), 871–874.

Carey, A. (1967). The Hawthorne studies: A radical criticism. *American Sociological Review*, *32*(3), 403–416.

Clore, G. L., & Huntsinger, J. R. (2007). How emotions inform judgment and regulate thoughts. *Trends in Cognitive Sciences*, *11*(9), 393–399.

Cole, D. A., & Preacher, K. J. (2014). Manifest variable path analysis: Potentially serious and misleading consequences due to uncorrected measurement error. *Psychological Methods*, *19*(2), 300–315.

Dasgupta, N., DeSteno, D. A., Williams, L., & Hunsinger, M. (2009). Fanning the flames of prejudice: The influence of specific incidental emotions on implicit prejudice. *Emotion (Washington, D.C.)*, *9*(4), 585–591.

Donaldson, S. I., & Grant-Vallone, E. J. (2002). Understanding self-report bias in organizational research. *Journal of Business and Psychology*, *17*(2), 245–260.

Downing, J. W., Judd, C., & Brauer, M. (1992). Effects of Repeated Expressions on Attitude Extremity. *Journal of Personality and Social Psychology*, 63, 17–29.

Dunn, J. R., & Schweitzer, M. E. (2005). Feeling and believing: The influence of emotion on trust. *Journal of Personality and Social Psychology*, *88*(5), 736–748.

Forster, J., & Dannenberg, L. (2010). GLOMO: A systems account of global versus local processing. *Psychological Inquiry*, 21, 175–197.

Forster, J., Friedman, R. S., & Liberman, N. (2004). Temporal construal effects on abstract and concrete thinking: Consequences for insight and creative cognition. *Journal of Personality and Social Psychology*, 87, 177–189.

Gawronski, B., & Hahn, A. (in press). Implicit measures: Procedures, use, and interpretation. In H. Blanton, & G. D. Webster (Eds.), *Foundations of social psychological measurement*. New York, NY: Taylor & Francis.

Godin, G., Sheeran, P., Conner, M., & Germain, M. (2008). Asking questions changes behavior: Mere measurement effects on frequency of blood donation. *Health Psychology*, *27*(2), 179.

Greenwald, A. G., Carnot, C. G., Beach, R., & Young, B. (1987). Increasing voting behavior by asking people if they expect to vote. *Journal of Applied Psychology*, *72*(2), 315–318.

Grice, J. W., Cohn, A., Ramsey, R. R., & Chaney, J. M. (2015). On muddled reasoning and mediation modeling. *Basic and Applied Social Psychology*, *37*(4), 214–225.

Gross, J. J. (2008). Emotion regulation. In M. Lewis, J. M. Haviland-Jones, & L. F. Barrett (Eds.), *Handbook of emotions* (3rd ed., pp. 497–512). New York, NY: Guilford Press.

Hunsinger, M., Isbell, L. M., & Clore, G. L. (2012). Sometimes happy people focus on the trees and sad people focus on the forest: Context-dependent effects of mood in impression formation. *Personality and Social Psychology Bulletin*, *38*(2), 220–232.

Huntsinger, J. R., Sinclair, S., Dunn, E., & Clore, G. L. (2010). Affective regulation of automatic stereotype activation: It's the (accessible) thought that counts. *Personality and Social Psychology Bulletin*, *36*(4), 564–577.

Iida, M., Shrout, P., Laurenceau, J. P., & Bolger, N. (2012). Using diary methods in psychological research. In H. Cooper (Ed.), *APA handbook of research methods in psychology: Foundations, planning, measures and psychometrics* (Vol. 1, pp. 277–305). Washington, DC: American Psychological Association.

Isbell, L. M. (2004). Not all happy people are lazy or stupid: Evidence of systematic processing in happy moods. *Journal of Experimental Social Psychology*, *40*(3), 341–349.

Kalokerinos, E. K., Greenaway, K. H., & Denson, T. F. (2015). Reappraisal but not suppression downregulates the experience of positive and negative emotion. *Emotion (Washington, D.C.)*, *15*(3), 271–275.

Kassam, K. S., & Mendes, W. B. (2013). The effects of measuring emotion: Physiological reactions to emotional situations depend on whether someone is asking. *PLoS One*, *8*, e64959.

Keltner, D., Locke, K. D., & Audrain, P. C. (1993). The influence of attributions on the relevance of negative feelings to personal satisfaction. *Personality and Social Psychology Bulletin*, *19*(1), 21–29.

Kidd, R. F. (1976). Manipulation checks: Advantage or disadvantage. *Representative Research in Social Psychology*, *7*, 160–165.

Kline, R. B. (2015). The mediation myth. *Basic and Applied Social Psychology*, *37*(4), 202–213.

Knowles, E. S., Coker, M. C., Scott, R. A., Cook, D. A., & Neville, J. W. (1996). Measurement-induced improvement in anxiety: Mean shifts with repeated assessment. *Journal of Personality and Social Psychology*, *71*(2), 352–363.

Krizan, Z., & Johar, O. (2012). Envy divides the two faces of narcissism. *Journal of Personality*, *80*(5), 1415–1451.

Landsberger, H. A. (1958). *Hawthorne revisited: management and the worker, its critics, and developments in human relations in industry*. Ithaca, NY: Cornell University.

Lench, H. C., Flores, S. A., & Bench, S. W. (2011). Discrete emotions predict changes in cognition, judgment, experience, behavior, and physiology: A meta-analysis of experimental emotion elicitations. *Psychological Bulletin*, *137*(5), 834–855.

Levitt, S. D., & List, J. A. (2011). Was there really a Hawthorne effect at the Hawthorne plant? An analysis of the original illumination experiments. *American Economic Journal: Applied Economics*, *3*(1), 224–238.

Lieberman, M. D., Inagaki, T. K., Tabibnia, G., & Crockett, M. J. (2011). Subjective responses to emotional stimuli during labeling, reappraisal, and distraction. *Emotion*, *11*(3), 468–480.

Lindsay, J. J., & Anderson, C. A. (2000). From antecedent conditions to violent actions: A general affective aggression model. *Personality and Social Psychology Bulletin*, *26*(5), 533–547.

Miller, D., & Form, W. (1951). *Industrial sociology*. New York, NY: Harper.

Morwitz, V. G., & Fitzsimons, G. J. (2004). The Mere-Measurement effect: Why does measuring intentions change actual behavior? *Journal of Consumer Psychology*, *14*, 64–74.

Morwitz, V. G., Johnson, E., & Schmittlein, D. (1993). Does measuring intent change behavior? *Journal of Consumer Research*, *20*, 46–61.

Moskowitz, D. S., Russell, J. J., Sadikaj, G., & Sutton, R. (2009). Measuring people intensively. *Canadian Psychology*, *50*(3), 131–140.

Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review*, *84*(3), 231–259.

Nolen-Hoeksema, S., Morrow, J., & Fredrickson, B. L. (1993). Response styles and the duration of episodes of depressed mood. *Journal of Abnormal Psychology*, *102*(1), 20–28.

Nyer, P. U. (2000). An investigation into whether complaining can cause increased consumer satisfaction. *Journal of Consumer Marketing*, *17*(1), 9–19.

Orne, M. T. (1962). On the social psychology of the psychological experiment: With particular reference to demand characteristics and their implications. *American Psychologist*, *17*(11), 776–783.

Ortner, C. N. M. (2015). Divergent effects of reappraisal and labeling internal affective feelings on subjective emotional experience. *Motivation and Emotion*, *39*(4), 563–570.

Podsakoff, P. M., MacKenzie, S. B., Lee, J.-Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology*, *88*(5), 879–903.

Powell, M. C., & Fazio, R. H. (1984). Attitude accessibility as a function of repeated attitudinal expression. *Personality and Social Psychology Bulletin*, *10*(1), 139–148.

Ray, R. D., McRae, K., Ochsner, K. N., & Gross, J. J. (2010). Cognitive reappraisal of negative affect: Converging evidence from EMG and self-report. *Emotion (Washington, D.C.)*, *10*(4), 587–592.

Ray, R. D., Wilhelm, F. H., & Gross, J. J. (2008). All in the mind's eye? Anger rumination and reappraisal. *Journal of Personality and Social Psychology*, *94*(1), 133–145.

Reis, H. T., & Gable, S. L. (2000). Event-sampling and other methods for studying everyday experience. In H. T. Reis & M. C. Judd (Eds.), *Handbook of research methods in social and personality psychology* (pp. 190–222). New York, NY: Cambridge University Press.

Riley, K. J., & Steitz, J. A. (2013). The "Observer Effect" in genome-wide surveys of protein-RNA interactions . *Molecular Cell*, *49*(4), 601–604.

Rusting, C. L. (1999). Interactive effects of personality and mood on emotion-congruent memory and judgment. *Journal of Personality and Social Psychology*, *77*(5), 1073–1086.

Salovey, P., Mayer, J. D., Goldman, S. L., Turvey, C., & Palfai, T. P. (1995). Emotional attention, clarity, and repair: Exploring emotional intelligence using the trait Meta-Mood Scale. In J. W. Pennebaker (Ed.), *Emotion, disclosure, and health* (pp. 125–154). Washington, DC: American Psychological Association.

Sandberg, T., & Conner, M. (2009). A mere measurement effect for anticipated regret: Impacts on cervical screening attendance. *British Journal of Social Psychology*, *48*(2), 221–236.

Sawyer, A. G., Lynch, J. G., & Brinberg, D. L. (1995). A Bayesian analysis of the information value of manipulation and con-founding checks in theory tests. *Journal of Consumer Research*, *21*(4), 581–595.

Scherer, K. R., Schorr, A., & Johnstone, T. (Eds.). (2001). *Appraisal processes in emotion: Theory, methods, research*. New York, NY: Oxford University Press.

Schmidt, F. L., & Hunter, J. E. (1996). Measurement error in psychological research: Lessons from 26 research scenarios. *Psychological Methods*, *1*(2), 199–223.

Sharpe, J. P., & Gilbert, D. G. (1998). Effects of repeated administration of the Beck Depression Inventory and other measures of negative mood states. *Personality and Individual Differences*, *24*(4), 457–463.

Sherman, S. J. (1980). On the self-erasing nature of errors of prediction. *Journal of Personality and Social Psychology*, *39*(2), 211–221.

Shrout, P. E., Stadler, G., Lane, S. P., McClure, M. J., Jackson, G. L., Clavél, F. D., … Bolger, N. (2018). Initial elevation bias in subjective reports. *Proceedings of the National Academy of Sciences*, *115*(1), E15.

Spector, P. E. (1994). Using self-report questionnaires in OB research: A comment on the use of a controversial method. *Journal of Organizational Behavior*, *15*(5), 385–392.

Tabibnia, G., Lieberman, M. D., & Craske, M. G. (2008). The lasting effect of words on feelings: Words may facilitate exposure effects to threatening images. *Emotion*, *8*, 307–317.

Tamir, M. (2009). What do people want to feel and why?: Pleasure and utility in emotion regulation. *Current Directions in Psychological Science*, *18*(2), 101–105.

Tate, C. U. (2015). On the overuse and misuse of mediation analysis: It may be a matter of timing. *Basic and Applied Social Psychology*, *37*(4), 235–246.

Thoemmes, F. (2015). Reversing arrows in mediation models does not distinguish plausible models. *Basic and Applied Social Psychology*, *37*(4), 226–234.

Thompson, R. J., Mata, J., Jaeggi, S. M., Buschkuehl, M., Jonides, J., & Gotlib, I. H. (2011). Concurrent and prospective relations between attention to emotion and affect intensity: An experience sampling study. *Emotion*, *11*(6), 1489–1494.

Trafimow, D. (2012). The role of mechanisms, integration, and unification in science and psychology. *Theory and Psychology*, *22*(5), 697–703.

Trafimow, D., & Rice, S. (2009). What if social scientists had reviewed great scientific works of the past? *Perspectives on Psychological Science*, *4*(1), 65–78.

Trapnell, P. D., & Campbell, J. D. (1999). Private self-consciousness and the five-factor model of personality: Distinguishing rumination from reflection. *Journal of Personality and Social Psychology*, *76*(2), 284–304.

Trope, Y., & Liberman, N. (2003). Temporal construal. *Psychological Review*, 110, 403–421.

Umashankar, N., Ward, M. K., & Dahl, D. W. (2017). The benefit of becoming friends: Complaining after service failures leads customers with strong ties to increase loyalty. *Journal of Marketing*, *81*(6), 79–98.

Urry, H. L. (2009). Using reappraisal to regulate unpleasant emotional episodes: Goals and timing matter. *Emotion*, *9*(6), 782–797.

Van Boven, L., Kane, J., McGraw, A. P., & Dale, J. (2010). Feeling close: Emotional intensity reduces perceived psychological distance. *Journal of Personality and Social Psychology*, *98*(6), 872–885.

Verduyn, P., & Brans, K. (2012). The relationship between extraversion, neuroticism and aspects of trait affect. *Personality and Individual Differences*, *52*(6), 664–669.

Volokhov, R., & Demaree, H. A. (2010). Spontaneous emotion regulation to positive and negative stimuli. *Brain and Cognition*, *73*(1), 1–6.

Wakslak, C. J., Trope, Y., Liberman, N., & Alony, R. (2006). Seeing the forest when entry is unlikely: Probability and the mental representation of events. *Journal of Experimental Psychology: General*, 135, 641–653.

Watkins, E. R. (2008). Constructive and unconstructive repetitive thought. *Psychological Bulletin*, *134*(2), 163–206.

Watson, D., & Clark, L. A. (1994). *The PANAS-X: manual for the positive and negative affect schedule – expanded form*. Iowa City, IA: Department of Psychology, University of Iowa.

Webb, T. L., Miles, E., & Sheeran, P. (2012). Dealing with feeling: A meta-analysis of the effectiveness of strategies derived from the process model of emotion regulation. *Psychological Bulletin*, *138*(4), 775–808.

Wheeler, L., & Reis, H. T. (1991). Self-recording of everyday life events: origins, types, and uses. *Journal of Personality*, *59*, 339–354.

White, K., & Van Boven, L. (2012). Immediacy bias in social-emotional comparisons. *Emotion (Washington, D.C.)*, *12*(4), 737–747.

Whitmer, A. J., & Gotlib, I. H. (2013). An attentional scope model of rumination. *Psychological Bulletin*, *139*(5), 1036–1061.