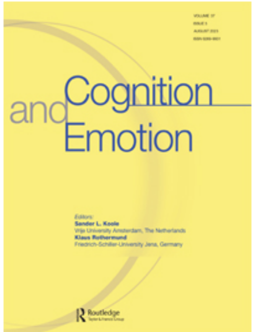

Cognition and Emotion

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/pcem20

Longitudinal and experimental investigations of implicit happiness and explicit fear of happiness

Amanda C. Collins, D. Gage Jordan, Gregory Bartoszek, Jenna Kilgore, Alisson N. S. Lass & E. Samuel Winer

To cite this article: Amanda C. Collins, D. Gage Jordan, Gregory Bartoszek, Jenna Kilgore, Alisson N. S. Lass & E. Samuel Winer (2023) Longitudinal and experimental investigations of implicit happiness and explicit fear of happiness, Cognition and Emotion, 37:5, 973-989, DOI: 10.1080/02699931.2023.2223907

To link to this article: https://doi.org/10.1080/02699931.2023.2223907

COGNITION AND EMOTION
2023, VOL. 37, NO. 5, 973–989
https://doi.org/10.1080/02699931.2023.2223907

Longitudinal and experimental investigations of implicit happiness and explicit fear of happiness

Amanda C. Collins [a,b,c], D. Gage Jordan [d], Gregory Bartoszek [e], Jenna Kilgore[a], Alisson N. S. Lass [f] and E. Samuel Winer [g]

[a]Department of Psychology, Mississippi State University, Mississippi State, MS, USA; [b]Center for Technology and Behavioral Health, Geisel School of Medicine, Dartmouth College, Lebanon, NH, USA; [c]Department of Psychiatry, Dartmouth-Hitchcock Medical Center, Lebanon, NH, USA; [d]Department of Psychology, Murray State University, Murray, KY, USA; [e]School of Psychology and Counseling, Fairleigh Dickinson University, Teaneck, NJ, USA; [f]Departments of Psychiatry and Psychology, Mayo Clinic, Rochester, MN, USA; [g]Department of Psychology, The New School for Social Research, New York, USA

ABSTRACT
Some individuals devalue positivity previously associated with negativity (Winer & Salem, 2016). Positive emotions (e.g. happiness) may be seen as threatening and result in active avoidance of future situations involving positivity. Although some self-report measures can capture emotions of happiness-averse individuals, they are not always capable of capturing automatic processing. Thus, we examined the association between implicitly-assessed happiness and explicit (i.e. self-reported) fear of happiness in three studies. In Study 1, participants completed the Fear of Happiness Scale (FHS) and an implicit measure of emotions at four-time points over approximately one year. The implicit measure required participants to choose which emotion (i.e. anger, fear, happiness, sadness, or none) best corresponded to 20 individual Chinese characters. In Studies 2 and 3, we utilized an experimental design, implementing a mood induction to emphasise the relationship between explicit fear of happiness and implicitly-assessed happiness. Participants completed the FHS and chose which emotion they believed the artist tried to convey in 20 abstract images. Results indicated that greater self-reported fear of happiness was related to reduced implicit happiness. Findings from these studies provide compound evidence that individuals who hold negative views of positivity may process implicit happiness in a devaluative manner.

ARTICLE HISTORY
Received 17 August 2022
Revised 16 May 2023
Accepted 4 June 2023

KEYWORDS
Reward devaluation theory; Happiness; Implicit measure; IMDES

Reward devaluation theory (RDT; Winer & Salem, 2016) posits that some individuals may avoid reward (e.g. positivity) as it may have been previously associated with negative outcomes. According to RDT, a subset of individuals who exhibit diminished approach and responsiveness patterns to rewarding stimuli may actually exhibit an *active*, automatic avoidance to reward, rather than a mere lack of approach toward reward or positivity (Winer & Salem, 2016). These individuals may have had past experiences where positive information that was initially seen as hopeful was met with disappointing or negative outcomes; thus, positivity, or the prospect of future positivity, is viewed as even more threatening than negative information. Support for this argument comes from meta-analytic findings providing evidence that depressed individuals exhibit a systematic bias away from positive information (Winer & Salem, 2016), which is otherwise not evident in psychologically-healthy individuals (Pool et al., 2016).

Fear of happiness

In line with the main tenets of RDT, research indicates that, for some individuals, positive emotions (e.g. happiness, joy, and hope) are feared, avoided, and

CONTACT Amanda C. Collins Amanda.C.Collins@dartmouth.edu Center for Technology and Behavioral Health, Dartmouth College, 46 Centerra Parkway Suite 300, Office # 338S, Lebanon, NH 03766
Supplemental data for this article can be accessed online at https://doi.org/10.1080/02699931.2023.2223907.

devalued. Some individuals, for example, believe that experiencing positive emotions makes them a bad person (Joshanloo & Weijers, 2014). Moreover, some individuals, including those in Eastern cultures, may not value happiness and view it as not worthy of pursuing. However, for most individuals who devalue positive emotions, happiness can be frightening due to previous negative experiences (Gilbert et al., 2012; Gilbert et al., 2014; Joshanloo et al., 2013; Joshanloo & Weijers, 2014). Thus, even the prospect of being happy may be viewed as fearful and result in negative emotions or other negative outcomes (Joshanloo & Weijers, 2014; Şar et al., 2019), thus resulting in some individuals avoiding or devaluing future positivity, in line with RDT. Individuals who fear positive emotions may believe that happiness is temporary or that only bad things happen when one is happy (Arieti & Bemporad, 1980). The Fear of Happiness Scale (FHS; Gilbert et al., 2012) is one self-report measure that examines one's negative views of happiness and positive emotions. Prior research has demonstrated that depressed individuals endorse more fear of happiness compared to controls and engage in other devaluative behaviours to decrease their experience of positive emotions, including dampening (Feldman et al., 2008; Gilbert et al., 2014; Gilbert et al., 2014; Joshanloo & Weijers, 2014; Quoidbach et al., 2010).

These findings are in line with the RDT framework, and the FHS may be a promising self-report measure that captures the devaluative processes of reward. Indeed, many items on the FHS reflect the devaluative processes of RDT (e.g. "I am frightened to let myself become too happy," "I feel I don't deserve to be happy," and "I worry that if I feel good something bad could happen"). Although the FHS is a relatively face-valid measure and includes many explicit statements likely made by hope – or happiness-averse individuals, RDT emphasises the *automaticity* of these biases. For example, the previous repeated pairings of happiness and disappointment may foster an association suggesting to the individual that any semblance of positivity will ultimately result in a negative outcome. As such, the explicit statements the individual endorses on the FHS can reflect these past aversive experiences and associations and may serve as a heuristic to avoid future positive events or activities (Gilbert et al., 2012). Therefore, the automatic processes that ultimately lead to these verbal, explicit statements regarding happiness may be best examined with "implicit" measures, which may not otherwise be evident when relying on self-report measures alone.

Use of an implicit measure

Prior work has also raised concerns about the limitations of self-report measures (Bartoszek & Cervone, 2022; Mauss & Robinson, 2009). Indeed, individuals may respond in a more socially-desirable manner such that they are more likely to report more positive emotions and less likely to report negative emotions. For example, individuals who are high in social desirability may be less likely to endorse a fear of happiness given its inherently negative emotionality. In addition, alexithymia has been found to be positively related with fear of happiness (Gilbert et al., 2012; Gilbert et al., 2014), and individuals with alexithymia may have difficulties identifying and describing their subjective emotional experiences with a self-report measure (Murphy et al., 2018). Thus, given the limitations of self-report measures, it is important to assess fear of happiness with other indices, including implicit measures.

Implicit measures are distinct from self-report (or "explicit") measures in that they assess mental contents or emotional states without requiring respondents to engage in introspective reporting (Gawronski, 2009). Because such measures can assess constructs automatically without relying on conscious deliberations (De Houwer et al., 2009), they can tap into processes and motivational goals that occur outside of conscious awareness (Tamir et al., 2013). Specifically, implicit measures can examine automatic processing of information, as measurement outcomes depend on unintentional responses, low effort, unconscious, and/or uncontrollable (De Houwer et al., 2009; Gawronski & Hahn, 2019). Further, depending on the measurement outcome, responses can also reflect a bias against (or for) certain stimuli (Gawronski & Hahn, 2019).

However, one criticism of implicit measures relates to their lack of specificity (among other criticisms, e.g. poor reliability; Brownstein et al., 2020; De Houwer & Moors, 2010), particularly with implicit emotion measures. Until recently, it was difficult, if not impossible, to assess emerging implicit fear of happiness. However, Bartoszek and Cervone (2017, 2022) have developed an implicit measure of emotions that captures the variability of emotional experience and differentiates distinct emotional states of the same valence. This measure thus provides the heightened level of specificity necessary to assess whether happiness, specifically, would be associated with explicit self-reported fear of happiness. Indeed, Bartoszek and Cervone (2017, 2022) have demonstrated that their implicit measure of emotions was able to capture the variability of emotional experience, as well as differentiate among emotions of the same valence (e.g. sad and angry). Moreover, their implicit measure demonstrated validity such that implicit emotions were related to different indices of emotion, including behavioural and physiological responses (Bartoszek & Cervone, 2022). As such, inclusion of Bartoszek and Cervone's (2017, 2022) implicit measure may allow us to better explain why some individuals end up fearing happiness or avoiding positivity as one's responses on the implicit measure may reflect their implicit emotions, which can influence their responding on a self-report measure.

For example, consider why happiness might be less likely to be implicitly endorsed. Low levels of implicitly assessed happiness could potentially indicate either (a) a mere lack of approach motivation toward positivity (i.e. a loss of motivation due to an expectancy that something is not desirable) or (b) an oppositional motivational tendency to avoid positivity (i.e. actively refraining from engaging with stimuli due to the expectation it can be harmful). However, if lack of happiness attributions indicated merely reduced approach motivation, one would not expect to find a robust relationship between happiness ratings and fear of happiness; instead, happiness would simply not have much of a relationship with negativity. Conversely, if there is an association between fear of happiness and implicitly measured happiness, this would provide evidence that individuals who associate happiness with fear may do so in part because of an automatic tendency to avoid positivity in general.[1]

Overview

No research to date has assessed whether differences in implicit endorsement of happiness are related to self-reported fear of happiness. Therefore, the present studies sought to better understand the relationship between these two constructs. To accomplish this, we examined these relationships in three separate studies: one incorporating a longitudinal assessment of these measures and two aiming to replicate these findings in cross-sectional experimental settings. In addition, consistent with our interest in assessing automatic implicit happiness, we sought to examine whether response time on the implicit measure may play a particularly important role in the first study.

Theoretical and empirical findings support the role of response time in assessing automatic processing with implicit measures. For example, the way in which emotion and cognition are linked together likely differ depending on whether respondents engage in slow and deliberate processing as opposed to quick, heuristic processing (Bartoszek & Cervone, 2022; Greifeneder et al., 2011; Siemer & Reisenzein, 1998). Regarding the latter type of processing, the feelings-as-information theory (Schwarz, 2011) states that respondents are more likely to use their current emotional state as a source of information when evaluating ambiguous stimuli (e.g. abstract characters or paintings). Conversely, slow and deliberate processing is less heuristic-based, requiring more analytic thought as a result (Palkovics & Takáč, 2016). This distinction between fast and slow processing is also made explicit in dual-processing theories of cognition, such that "Type I" processes are rapid and autonomous, whereas "Type II" processes involve higher-order thought and reasoning (e.g. working memory functioning; Evans & Stanovich, 2013).

Further, Bartoszek and Cervone's (2017, 2022) findings are consistent with these theoretical accounts. After participants underwent mood inductions in the experiments presented in these two papers, the effect of these inductions on ratings of ambiguous images were more pronounced among fast-, compared to slow-responding, participants, suggesting that these ratings of these images relied on the more heuristic or Type I processes (whereas slow-responding participants likely examined the context of the images using analytical Type II processes). These results suggest that ratings of these images relied on the more heuristic processes and thus set precedence for examining the role of response time in the implicit measure of emotions detailed further below. However, to our knowledge, the extent to which automatic emotional processing (of ambiguous stimuli) is associated with fear of happiness has yet to be examined. As such, our guiding framework (i.e. RDT) and associations between implicit measures of emotion and fear of happiness are examined in the following studies.

Study 1

In Study 1, we investigated the relationship between implicit happiness and explicit fear of happiness longitudinally. Drawing from the reward devaluation framework (Winer & Salem, 2016), we hypothesised that reduced happiness attributions on the implicit measure would be associated with higher levels of self-reported fear of happiness over time. Specifically, we predicted that implicit happiness, which represents one's unconscious emotions, would be associated with higher levels of explicit happiness. In addition, we hypothesised that this relationship would interact with response time on the implicit measure, such that faster respondents would be more likely to evidence this effect (Bartoszek & Cervone, 2017, 2022). Lastly, we included implicit endorsements of sadness as a comparison to implicit endorsements of happiness in analyses to assess the measure's discriminant validity. We predicted that the above relationships would be present only with people with reduced endorsements of happiness, consistent with our theory-driven belief that the reduction of implicit happiness is specifically related to a fear of happiness (Winer & Salem, 2016).

Method

Participants

Participants were English-speaking U.S. residents who completed a battery of online questionnaires at four separate time points spanning approximately one calendar year (79% White, 6.6% Black, 4.4%, Hispanic, 3.8% Asian or Pacific Islander, 1.0% American Indian or Alaskan Native, and 2.1% "Other," at the initial time point). One thousand seven participants initially completed the full battery of questionnaires. There were 705 validly-responding participants at Time 1 (501 females, $M_{age} = 37.80$, age range: 18–77 years), 375 participants at Time 2 (266 females) 292 participants at Time 3 (209 females), and 220 participants (153 females) at Time 4. Participants completed measurements at Time 2 approximately one month after Time 1 ($M_{days} = 31.81$, $SD = 6.99$ days), at Time 3 approximately three months after Time 1 ($M_{days} = 97.81$, $SD = 25.36$ days), and at Time 4 approximately 11 months after Time 1 ($M_{days} = 319.99$, $SD = 24.65$ days).

Procedure

Participants were obtained from a larger longitudinal study (see Jordan et al., 2018) completed via Amazon.com's Mechanical Turk (MTurk). Prior research indicates that MTurk is an effective and reliable method for collecting data regarding clinical and affective variables (Chandler & Shapiro, 2016; Shapiro et al., 2013). This study was approved by the university's Institutional Review Board (IRB #14-196). Participants who completed the implicit measure and Fear of Happiness Scale, as well as those who correctly responded to a validity item at each time point (see below), were included in the analyses.[2] Participants were paid $1 at the end of Time 1, 2, and 3, and $3 at the end of Time 4 for participation, initially consenting to complete the survey and to be re-contacted for future waves before continuing to the actual measures. Participants were re-contacted using their de-identified Mturk ID via python and R (Leeper, 2014). At the end of the battery for each wave, a self-satisficing question ensured that participants responded validly. The validity item consisted of a paragraph ostensibly discussing emotions but in actuality instructing participants to ignore the main question, to choose the "other" response option, and to type, "I've read the instructions" in the text field. If a participant did not consent to be re-contacted or successfully answer the validity question at any wave, they were excluded. Thus, the number of participants at each wave described above are specific only to participants who correctly answered the validity item each time and consented to be re-contacted for further study.

Measures

Implicit measure of emotions. The implicit measure of emotions used in Study 1 was previously used by Bryant et al. (2017) and based on the early efforts by Bartoszek (2017) and Bartoszek and Cervone (2017) to develop an implicit measure of distinct emotional states. Participants viewed and rated 20 Chinese characters (i.e. ideograms), presented one at a time, and were asked to indicate which emotion best corresponded to each character (for further description of the measure, see Bartoszek, 2017; Bartoszek & Cervone, 2017). The emotion answer choices (i.e. anger, fear, happiness, sadness, or none) were presented as a nominal scale and were listed in a single line below of and at the same time as the Chinese character. Participants were instructed to enter the number on their keyboard corresponding to the order of the emotion within the line (e.g. entering "1" indicated anger as it was presented first in the line of emotions choices). Each character remained on the screen until the participant selected an emotion. The instructions asked participants to make ratings based on their first impression of each picture and not to spend too much time evaluating the character. Happiness and sadness attributions, along with response time from this measure were used in the main analyses described below. Of note, frequency of happiness and sadness attributions were used to assess implicit emotional states on this measure. As such, standard internal consistency values (e.g. Cronbach's alpha) are not provided, as the nature of the scale differs from that of typical self-report measures.

Fear of Happiness Scale (FHS). The FHS (Gilbert et al., 2012) is a nine-item Likert scale that assesses one's perception and apprehension around feeling happy and experiencing positive feelings in general (e.g. "I feel I don't deserve to be happy"). Response options range from 0 ("not at all like me") to 4 ("extremely like me"). Higher scores indicate greater fear of happiness, and the mean and standard deviation of the FHS in the current study at Time 4 ($M = 11.58$, $SD = 9.45$) is comparable to previous studies (Gilbert et al., 2012; Gilbert et al., 2014). The FHS demonstrated excellent internal consistency at each time point in this study (α range: .93 - .94), consistent with previous research (Gilbert et al., 2012; Gilbert et al., 2014).

Data analysis

We sought to investigate the effects of happiness and sadness attributions from the implicit emotion measure on self-report fear of happiness via a linear mixed (or "multilevel") model. Linear mixed models in the context of longitudinal are particularly robust compared to other procedures, such as repeated-measures analysis of variance (rmANOVA; Hox et al., 2018). These analyses were conducted using the R package *nlme* (Pinheiro et al., 2018). A full, detailed write-up of multilevel parameter selection, model-fitting procedures, and missing data analysis are provided as Supplemental Materials for brevity. In sum, these additional checks and analyses detailed in the Supplemental Materials suggest the full multilevel models were adequately estimated. Outside of the additional parameters included within multilevel models, these models can be interpreted in a similar manner as ordinary least squares (OLS) regression models.

Two separate full linear mixed models were analyzed: (a) one using aggregate response time on the implicit measure of emotions, and (b) one using median response time to categorise participants into "fast" or "slow" respondents. Participants were categorised as fast if their median response time on the implicit measure of emotions was below 5.00 s and slow if their median response time was above 5.00 s. The decision to use a median split and 5.00 s as the threshold for categorising fast and slow respondents was based on prior findings using an implicit measure of emotions similar to the measure used in this study (Bartoszek & Cervone, 2017, 2022).

Individual response time trials under 2.00 s were excluded on the basis that these trials may indicate random responses, taking precedence from prior research (Bartoszek & Cervone, 2022). Across all time points and trials, this exclusion accounted for less than 1% of all response time data.

Results & discussion

See Table 1 for repeated-measures correlations between variables (Bakdash & Marusich, 2017). Correlations between the variables used in the full multilevel models at each time point are provided as supplemental materials (Tables S1-S4). See Table 2 for an overview of the linear mixed models from this study, with the following paragraph providing a more streamlined description of the results from the multilevel models.[3] See the supplemental materials for a full overview of the models' parameters and estimation procedures. Here, we report and discuss the main effects of the full multilevel models in a similar manner to a regression model.

Overall, both models (i.e. the full models including response time) support the hypothesis that reduced implicitly measured happiness is associated with greater self-reported fear of happiness. For example, in the first model (using aggregate response time), we found a significant main effect of happiness

Table 1. Repeated Measures Correlations Between Variables in Study 1.

Variable	1.	2.	3.
1. Happiness Attributions	-		
2. Sadness Attributions	−0.19*	-	
3. Fear of Happiness Scale (FHS)	−0.06†	0.04	-
4. Response Time (Aggregate)	0.04	0.01	−0.02

Note. * $p < .001$. Repeated measures correlations were computed using the R package *rmcorr* (Bakdash & Marusich, 2017). †This correlation, however, suggests a trend ($p = 0.08$), with the supplemental tables showing that happiness attributions and FHS scores were significantly correlated at each time point.

Table 2. Multilevel Model Results Predicting Fear of Happiness Scores in Study 1.

	Fear of Happiness					
	Unconditional Model		Random Slopes Model		Multilevel Model	
Predictors	*Unstandardized Coefficient*	*SE*	*Unstandardized Coefficient*	*SE*	*Unstandardized Coefficient*	*SE*
Level 2						
Happiness Attributions					−0.07**	0.02
Sadness Attributions					0.03	0.02
Response Time (RT)†					−0.01	0.02
Happiness Attributions x RT†					0.01	0.01
Time			0.001	0.01	−0.01	0.01
Level 1						
Intercept	0.90	0.46	0.90	0.46	0.88	0.47
BIC	3606.23		3625.38		3545.54	

Note. ** $p < .001$. The intercept reflects the random intercepts portion of each model, where standard error values reflect the residual component of this random effect. All models were constructed with maximum likelihood estimation. † RT analyzed as a median split in a separate model also showed non-significant effects.

attributions, $b = -.07$, $p < .001$, 95% confidence interval (CI) [−0.11, −0.03], suggesting that fewer happiness attributions on the implicit emotion measure were associated with greater self-reported fear of happiness scores. However, the main effect of sadness attributions, $b = .03$, $p = 0.08$, 95% CI [−0.01, 0.07], response time, $b = -.001$, $p = 0.81$, 95% CI [−0.03, 0.03], and the interaction between response time and happiness attributions, $b = .001$, $p = 0.99$, 95% CI [−0.02, 0.02], were not significant. Similarly, in our next model (with the response time variable coded as a median split), we again found a significant main effect of happiness attributions, $b = -.07$, $p < .001$, 95% CI [−0.11, −0.03]. In addition, the main effect of sadness attributions, $b = .03$, $p = 0.08$, 95% CI [−0.01, 0.07], response time, $b = -.015$, $p = 0.49$, 95% CI [−0.06, 0.03], and the interaction between response time and happiness attributions, $b = -.017$, $p = 0.21$, 95% CI [−0.04, 0.01], were again not significant. As such, response time overall appeared to play a negligible role in these analyses and thus did not support the hypothesis that response time would interact with happiness attributions.

Study 2

The findings from Study 1 provided initial evidence of a relationship between reduced implicit happiness and greater self-reported fear of happiness. The non-experimental design of Study 1 might have not fully captured the devaluative processes of happiness that may occur in response to momentary positivity; thus, Study 2 utilised an experimental design to extend this investigation of the relationship between implicit happiness and explicit fear of happiness. Bartoszek and Cervone (2017) included emotion inductions to induce discrete emotions (e.g. sadness), and found that individuals who received this induction endorsed the respective implicit emotion (e.g. sadness) more frequently than individuals who received a different induction (e.g. relaxation).

Thus, the purpose of Study 2 was to include a positive mood induction procedure (MIP) to investigate how fear of happiness relates to implicit happiness following a positive MIP. We predicted that individuals, in general, who received the positive MIP would endorse implicit happiness more after the MIP compared to a) their endorsement of implicit happiness prior to the MIP and b) individuals who received the neutral MIP. We also predicted that, in contrast, implicit happiness would be endorsed *less* for individuals who fear happiness and received the positive MIP given their devaluative views of happiness.

Although the version of the implicit measure used in Study 1 was previously used and validated (see Bryant et al., 2017), we did not assess for whether participants in Study 1 were familiar with or spoke the Chinese language. In addition, recent research has shown the Implicit Measure of Distinct Emotional States (IMDES), an implicit measure which utilises abstract paintings, to be predictive of self-report, behavioural, and physiological indices of emotional experience (Bartoszek & Cervone, 2017, 2022). Several studies revealed that the convergent, discriminant, criterion, and incremental validity of the IMDES was enhanced because of its one core feature: the time constraint for responding. This is consistent with previous work suggesting that when responding slowly, participants are more likely to rely on information other than their emotional states (e.g. "cold" cognitive processes; Forgas, 1995). Conversely, if given a limited timeframe to respond (i.e. responding under time pressure), participants use automatic, emotion-based processes in evaluative judgment (Bartoszek & Cervone, 2022; De Houwer et al., 2009). Thus, although we did not find an effect of response time when examining associations between implicit happiness and fear of happiness in Study 1, given the goal of Study 2 to experimentally manipulate happiness, we employed the IMDES to index automatic emotion processes more accurately.

Method

Participants

Participants ($N = 255$) were recruited from the psychology subject pool at a large southern university (157 females, $M_{age} = 19.20$, age range: 18-50, 67.5% White, 22% Black, 3.6% Hispanic, 2.8% Asian or Pacific Islander, 1.2% American Indian or Alaskan Native, and 1.2% "Other"). They were randomly assigned to either the positive ($n = 144$) or neutral mood-induction ($n = 111$) condition (MIP). Four (1.57%) participants were not included in the analyses due to apparatus error or participant non-compliance (e.g. using their cellphones during the experiment).

Mood Induction Procedure (MIP)

Each MIP condition involved a two-minute video, in which a narrator described an ostensible study that participants could be invited to later in the semester. To conceal the purpose of the video, the experimenter informed participants that the goal of the video was to gauge their interest in participating in the future study. They were instructed to visualise themselves participating in this study, whose description differed for the positive and neutral inductions. Participants in the positive MIP were told that they would have the opportunity to work on increasing their overall level of positive emotions in the future study. This film described the various activities (e.g. gratitude, mindfulness) that they would engage in during the prospective study to increase their positive emotions. Participants in the neutral MIP were told that they would have the opportunity to work on maintaining their current level of emotions. This film described the various activities that they would engage in during the future study to maintain the stability of their current emotions. Twelve (4.7%) participants were not included in the analyses due to correctly guessing the purpose of the study or mood induction.

Prior to the current study, we conducted a separate pilot study to examine the effectiveness in eliciting positive and neutral moods ($N = 34$). We developed a questionnaire to assess the valence and arousal dimensions of the circumplex model of affect (Russell, 1980, 2003) as well as the liking of the MIP videos. Compared to individuals in the neutral mood induction ($n = 17$), participants in the positive mood induction ($n = 17$) rated the video as being more pleasant, $t(32) = -2.85$, $p = .008$, $d = .98$, but did not differ in their ratings for arousal and liking of the two videos. In addition, participants in the positive MIP endorsed the video as eliciting happiness more than all other emotions (e.g. anger, fear, sadness, and no emotion). Thus, albeit from a limited sample size, our findings indicate that there were differences in the valence (e.g. pleasantness) between the positive and neutral MIPs. We did not include these face-valid questions regarding pleasantness and emotion elicited in the current study in an effort to keep the intention of our hypotheses concealed and reduce demand effects.[4]

Measures

Implicit Measure of Distinct Emotional States (IMDES). The implicit measure of emotions was similar to the one used in Study 1 with three exceptions to make it more aligned with the IMDES developed by Bartoszek and Cervone (2022). First, Chinese characters were replaced by 30 abstract images. Second, participants had up to 5 s to rate each image and were instructed to select an answer on the screen in the last 3 s. Imposing time pressure has been shown to make the measure more sensitive to variations in emotional states. Third, the first five images were treated as practice trials and are not included in analyses, and the last five images were excluded from analyses to reduce practice effects, resulting in 20 experimental trials. Following prior analytic approaches, we did not include the choice "none" in the current analysis (Bartoszek & Cervone, 2017, 2022).[5] This implicit measure has been successfully used in prior research to assess the effectiveness of MIPs (Bryant et al., 2017) and examine implicit emotions following a MIP (Bartoszek & Cervone, 2017; Bartoszek & Cervone, 2022).

Fear of Happiness Scale (FHS). The version of the FHS used in Study 2 was the same as the version used in Study 1. The mean and standard deviation of the FHS in Study 2 ($M = 7.70$, $SD = 6.28$) is comparable to Study 1 and previous studies (Gilbert et al., 2012; Gilbert et al., 2014). The FHS in this study also demonstrated good internal consistency (α = .85).

Procedure

This study was approved by the university's Institutional Review Board (IRB #18-370). The IMDES was administered both before and after the MIP. After the second administration of the IMDES, participants completed the FHS. They also completed a verbal, funnelled debrief with a trained researcher at the end of the session to assess for demand effects. These questions included asking participants about their general thoughts about the study, what they believed the purpose of the study was, and what they believed the purpose of the video was.

Results and discussion

Count scores for the IMDES were created for trials that were completed both before and after the MIP by counting the number of times participants recorded a response for each emotion, resulting in two count scores (i.e. pre- and post-count, respectively) for each emotion. The frequency that each of the four emotions (fear, anger, happiness, and sadness) was selected before and after the mood induction was calculated by multiplying the total possible responses (i.e. 20) and then dividing the count of each emotion by the total number of valid responses for both pre- and post-MIP (see equation below), resulting in two frequency scores for each emotion to be used in our analyses.

The skewness and kurtosis values for all pre- and post-MIP emotions were less than 1 (Kline, 2015). Eleven (4.3%) participants were not included in the analyses due to having insufficient IMDES data (i.e. missing responses for more than 5 images for pre-MIP or post-MIP; Bartoszek & Cervone, 2022).

Data from 228 participants were included for analyses following the exclusion criteria discussed above. One-hundred and twenty-six ($n = 126$) were included in the positive MIP, and 102 participants were included in the neutral MIP. We first examined the FHS scores between the two MIP conditions. An independent t-test indicated that participants in the positive MIP ($M = 8.57$, $SD = 6.45$) had higher FHS scores than participants in the neutral MIP ($M = 6.64$, $SD = 5.92$), $t(224) = -2.33$, $p = .02$, 95% CI [−3.57, −0.30], Cohen's $d = 0.31$; however, the difference was of a small effect size.[6]

We conducted a three-way mixed ANOVA to examine the effect of MIP (neutral, positive) and assessment period (pre-MIP, post-MIP) on implicitly assessed emotions (happiness, sadness, anger, fear). MIP condition was included as the between-subject factor, whereas assessment period and emotion were included as within-subject factors.[7] Results showed a significant main effect of emotion, $F(3, 224) = 38.92$, $p < .001$, $\eta_p^2 = .34$. Simple contrasts revealed that participants endorsed more images as conveying happiness than all other emotions at both pre- and post-MIP (happiness: $M = 5.612$; anger: $M = 3.878$; fear: $M = 4.135$; sadness: $M = 4.290$; happiness vs anger: $F(1, 226) = 100.87$, $p < .001$, $\eta_p^2 = .31$; happiness vs fear: $F(1, 226) = 99.09$, $p < .001$, $\eta_p^2 = .31$; happiness vs sadness: $F(1, 226) = 72.50$, $p < .001$, $\eta_p^2 = .24$).

Neither the main effect of assessment period, $F(1, 226) = 1.030$, $p = .31$, $\eta_p^2 = .01$, nor the interactions were significant (assessment period x MIP: $F(1, 226) = .01$, $p = .94$, $\eta_p^2 < .001$; emotion x MIP: $F(3, 224) = 1.77$, $p = .15$, $\eta_p^2 = .02$; assessment period x emotion: $F(3, 224) = 0.88$, $p = .45$, $\eta_p^2 = .01$; assessment period x MIP x emotion: $F(3, 224) = 0.50$, $p = .68$, $\eta_p^2 = .01$). These results suggest a lack of MIP effect such that there were no differences in the endorsement of each emotion on the IMDES between the pre-manipulation and post-manipulation assessments for either mood induction.

To examine how FHS scores may influence these findings, we also conducted a mixed ANCOVA with FHS scores as the covariate. Including FHS in the model did not alter our findings. Results suggest that there was only a main effect of emotion, $F(3, 220) = 22.36$, $p < .001$, $\eta_p^2 = .23$. There were no main effects of time or assessment or any significant interactions. However, a trend of an interaction of time x FHS scores emerged, $F(1, 222) = 3.39$, $p = .07$, $\eta_p^2 = .02$. The main effect of emotion was further examined with simple contrasts and revealed similar findings as those in the ANOVA above. See Table 3 for the correlations between variables used in this study.

Table 3. Intercorrelations of Average IMDES Variables and FHS in Study 2.

Measure	1.	2.	3.	4.	5.
1. FHS	-				
2. Anger	-.01	-			
3. Fear	.02	.08	-		
4. Sadness	.05	-.09	-.09	-	
5. Happiness	-.14*	-.32***	-.08	-.21**	-

Note. $N = 226$. *$p < .05$, **$p < .01$, ***$p < .001$

Given the non-significant assessment period by MIP interaction, we also examined how FHS scores relate to the average emotional response. To this end, we averaged the pre- and post-MIP frequencies for each emotion. We regressed FHS scores on the average emotions measured via the IMDES (Table 4). We included anger, fear, and sadness as predictors in the first step and happiness as a predictor in the second step. FHS was the dependent variable. The three emotions in Step 1 accounted for only 0.3% of the variance in FHS scores. The overall model significant was not significant, nor did any of the three emotions independently predict FHS scores. With happiness in the model (Step 2), the emotions accounted for 2.5% of the variance in FHS scores. Happiness ($\beta = -.16$, $p = .03$) significantly predicted FHS scores, suggesting that participants with greater FHS scores endorsed fewer images as conveying happiness. This finding was discriminant as none of the other three emotions were significantly related to FHS scores.

Our results suggest that the significant differences between MIP groups for FHS scores may have affected these null findings as participants in the positive MIP group endorsed greater levels of FHS scores. Whereas we predicted that participants in the positive MIP would endorse greater images as conveying happiness after the mood induction, it is possible that the higher levels of FHS scores in the positive MIP group resulted in a *decrease* of endorsement of images conveying happiness. In addition, research suggests that there may be a positivity offset such that participants are already experiencing slightly higher levels of positive affect before the induction (Cacioppo & Berntson, 1994), so informing participants in the neutral mood induction that they would maintain their current level of emotions may have resulted in them maintaining or increasing their prior levels of positive affect. As a result, this positivity offset may have contributed to the overall reduced endorsement of happiness from pre- to post-mood induction. Thus, replication of this study is warranted to investigate our findings further and determine the effectiveness of the mood induction in another, larger sample without a positivity offset.

Table 4. Hierarchical Multiple Regression Analyses Predicting FHS from IMDES Emotions in Study 2.

Predictor	ΔR^2	β
Step 1	.003	
Anger		-.01
Fear		.03
Sadness		.05
Step 2	.022*	
Anger		-.06
Fear		.01
Sadness		.01
Happiness		-.16*

Note. $N = 226$. $*p < .05$

Study 3

The findings from Study 2 showed reduced happiness among individuals who fear happiness (i.e. those with higher FHS scores). However, there were a few limitations of Study 2, as noted above, which may have made it difficult to draw conclusions about the lack of differences between the MIP groups for each emotion chosen on the IMDES given that individuals who fear happiness may respond differently to the positive MIP.

Study 3 was a follow-up study that improved upon Study 2 in several significant ways. First, we pre-registered this study online via aspredicted.org (https://aspredicted.org/7cy3w.pdf) before conducting analyses. Second, we recruited a larger sample size to investigate the effectiveness of the MIP. Third, we included practice trials for the IMDES to familiarise participants with the task prior to beginning the experimental trials. As outlined in our pre-registration, we had three hypotheses, stemming from our findings in Study 2. First, we hypothesised a significant interaction between MIP x emotion in the mixed ANOVA such that participants in the positive MIP would endorse more images as conveying happiness than participants in the neutral MIP. Second, we hypothesised that there would be a main effect of emotion in the mixed ANOVA such that participants would endorse more images as conveying happiness than all other emotions, replicating findings from Study 2. Third, we hypothesised that this finding would be the opposite in a hierarchical regression such that fear of happiness would be inversely related to implicit happiness, also replicating findings from Study 2.

Method

Participants

Eight-hundred and eighty-nine ($N = 889$) participants were initially recruited from the psychology subject pool at a large southern university (587 females, $M_{age} = 19.37$, age range: 18-76, 74.5% White, 16.1%

Black, 2.6% Hispanic, 2.6% Asian or Pacific Islander, 0.6% American Indian or Alaskan Native, 1.1% "Other", and 2.3% indicated more than one race). To ensure that participants who participated in Study 2 did not participate in the current study, we applied a filter to the current study's listing in the psychology subject pool so those who participated in Study 2 were unable to see the listing or sign up for the current study. Participants were randomly assigned to either the positive ($n = 447$) or neutral MIP ($n = 442$) condition.

Mood induction procedure

The MIP remained the same as in Study 2; however, this study was completed online via Qualtrics.[8] Participants were randomised to each MIP using the randomiser function in Qualtrics' survey flow. We also recorded how long participants watched the video and excluded participants who watched the video for less than 2 min (i.e. the length of the video) to ensure that participants engaged with the entire mood induction procedure. Two-hundred and seventy-one (30.5%) participants did not watch the entire MIP video.

In contrast to the methodology of Study 2, we did include the face valid questions assessing pleasantness and emotion elicited by the MIP in the current study in an effort to better understand the effects of the MIP. Our findings indicate that the MIP did influence participant responding on this questionnaire such that there was both a trend regarding valence with participants in the positive MIP indicating greater valence of the video, $F(1, 524) = 3.75$, $p = .053$, $\eta_p^2 = .007$, observed power = 0.49 and participants in the positive MIP indicating greater arousal of the video, $F(1, 524) = 11.88$, $p = .001$, $\eta_p^2 = .022$, observed power = 0.93. In addition, participants in the positive MIP strongly endorsed the video as eliciting happiness more than all other emotions and more than individuals in the neutral MIP, $X^2 = 70.33$, $p < 0.001$, Cramer's V = 0.37.

Measures

Implicit Measure of Distinct Emotional States (IMDES). We utilised the IMDES as described in Study 2 with additional two changes to match the final version recommended in Bartoszek and Cervone (2022). First, unlike in our Study 2, participants could not select an answer in the first 2 s. Second, participants responded to four practice trials that were not scored, which were not included in Study 2, to become familiar with the task. Third, participants were given feedback (i.e. "try to respond faster") if they did not respond within the 5 s to the last three practice images. After the practice images, participants viewed and rated 20 abstract images that served as the experimental trials.

Scoring and analyses of the IMDES data were identical to those in Study 2. The skewness and kurtosis values for all pre- and post-MIP emotions were less than 1 (Kline, 2015). Thirty-two (3.6%) participants did not have sufficient IMDES data (i.e. having missing responses for more than five images for pre-MIP or post-MIP).

Fear of Happiness Scale (FHS). The version of the FHS used in Study 3 was the same as the version used in Studies 1 and 2. The mean and standard deviation of the FHS for the overall sample in the current study ($M = 10.32$, $SD = 8.48$) is comparable to Studies 1 and 2 and previous studies (Gilbert et al., 2012; Gilbert et al., 2014). The FHS in this study also demonstrated good internal consistency ($\alpha = .92$).

Procedure

This study was approved by the university's Institutional Review Board (IRB #19-554). Participants completed the MIP, IMDES before and after the MIP, and the FHS online. We included the same validity item as Study 1 to assess for valid responding in this study, and participants also answered the same questions used in Study 2 for a funnelled debrief at the end of the session to assess for demand effects. They were instructed to type their answers into the respective textboxes. One-hundred and thirty (14.6%) participants failed the validity item, and 52 (5.8%) participants correctly guessed the purpose of the study or MIP.

Results and discussion

Three hundred and seventy-eight (42.52%) participants were excluded due to meeting criteria for at least one of the exclusion criteria discussed above. Thus, data from 511 participants were included in the current analyses. Two-hundred and sixty-three participants ($n = 263$) were included in the neutral MIP, and 248 participants were included in the positive MIP. Results from a t-test indicated that there were no significant differences in FHS scores between participants in the two MIP groups (Positive MIP: $M = 9.81$, $SD = 7.95$; Neutral MIP: $M = 10.81$, $SD =$

8.94), $t(509) = -1.33$, $p = .19$, 95% CI [−0.48, 2.47], Cohen's $d = 0.12$.

To examine the effect of MIP on the endorsement of emotions, we conducted a three-way mixed ANOVA. MIP was included as the between-subject factor, whereas assessment period (pre-MIP, post-MIP) and emotion were included as within-subject factors.[9] Results suggest that there was a main effect of emotion, $F(3, 507) = 18.83$, $p < .001$, $\eta_p^2 = .10$, replicating findings from Study 2. Simple contrasts revealed that participants endorsed more images as conveying happiness than all other emotions, similar to findings in Study 2 and providing support for our first hypothesis (happiness: $M = 4.898$; anger: $M = 4.541$; fear: $M = 4.260$; sadness: $M = 4.134$; happiness vs anger: $F(1, 509) = 7.87$, $p = .005$, $\eta_p^2 = .02$; happiness vs fear: $F(1, 509) = 30.29$, $p < .001$, $\eta_p^2 = .06$; happiness vs sadness: $F(1, 509) = 43.59$, $p < .001$, $\eta_p^2 = .08$). A main effect of assessment period, $F(1, 509) = 4.55$, $p = .03$, $\eta_p^2 = .009$, also emerged. Results also revealed significant interactions of assessment period x emotion, $F(3, 507) = 4.15$, $p = .006$, $\eta_p^2 = .025$, and assessment period x emotion x MIP, $F(3, 507) = 7.13$, $p < .001$, $\eta_p^2 = .04$ (see Figure 1). These results suggest that the MIP was effective as evidenced by the full factorial interaction in the ANOVA: implicit happiness was endorsed more at the post-manipulation assessment for the positive MIP only (positive pre-MIP: 4.679; positive post-MIP: 5.404; neutral pre-MIP: 4.846; neutral post-MIP: 4.664), providing support for our third hypothesis.

We also conducted a mixed ANCOVA with FHS scores as the covariate. Results suggest that there was only a main effect of emotion, $F(3, 505) = 10.84$, $p < .001$, $\eta_p^2 = .06$. There was no main effect of assessment or any significant interactions. However, a marginally significant interaction of emotion x FHS scores emerged, $F(3, 505) = 2.16$, $p = .09$, $\eta_p^2 = .01$. The main effect of emotion was further examined with simple contrasts and revealed similar findings as those in the ANOVA above. See Table 5 for the correlations between variables used in this study.

To investigate our second hypothesis, we again averaged the pre- and post-MIP frequencies for each emotion and regressed FHS scores on the average emotions measured via the IMDES. The results of the hierarchical regression are presented in Table 6 and indicate that implicit anger, not happiness, predicted FHS scores, contradicting our hypothesis. We included anger, fear, and sadness as independent variables in the first step and happiness as an independent variable in the second step. Fear of happiness was the dependent variable. The three emotions in Step 1 accounted for 1.5% of the variance in FHS scores. The overall model was significant, and anger independently predicted FHS scores ($\beta = .11$, $p = .02$). The emotions in Step 2 accounted for 1.7% of the variance in FHS scores. Happiness ($\beta = -.05$, $p = .33$) did not significantly predict FHS scores, but anger again independently predicted FHS scores ($\beta = .10$, $p = .04$). Fear and sadness again did not independently predict FHS scores.

Results from the hierarchical regression did not support our third hypothesis that happiness would be implicitly endorsed less by individuals who scored higher on the FHS. Instead, our results suggested that there was a positive relationship between anger and FHS scores such that anger was endorsed more by individuals who scored higher on the FHS; however, we still observed a negative non-significant relationship between happiness and FHS scores. This is indeed an intriguing finding. Fear and anger demonstrate higher levels of arousal, as evidenced by the circumplex model of emotions (Russell, 1980; 2003). Moreover, it is plausible that fear was induced through the positive MIP (for those who fear happiness), resulting in participants also experiencing anger in the moment. Thus, our findings may suggest that informing participants who fear happiness that they could learn how to increase their overall happiness or maintain their current mood could create a more intense, negative emotional response, such as anger, for those who devalue positivity. For example, telling someone who already has negative views of happiness that they can learn to be happier may evoke an aggressive response, which we may not have seen in Study 2 given that participants were already experiencing high levels of happiness before the MIP.

General discussion

In the three present studies, we examined the relationship between implicit happiness and explicit fear of happiness. We hypothesised that reduced implicit happiness, as measured by two implicit measures (Bartoszek & Cervone, 2017, 2022), would be associated with greater explicit fear of happiness, as measured by the FHS (Gilbert et al., 2012). We found longitudinal evidence in Study 1 to support this hypothesis: fewer happiness attributions on the implicit measure predicted greater fear of happiness

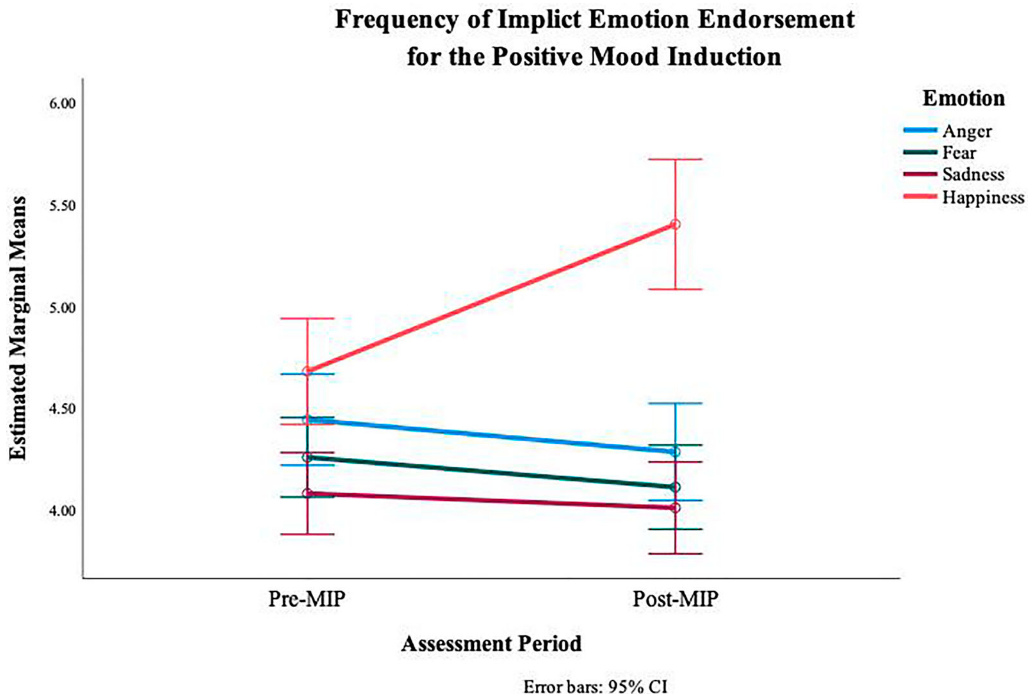

Figure 1. Results of the mixed ANOVA demonstrating the change of implicit emotion endorsement from pre- to post-induction for the positive mood induction procedure for Study 3.

scores. We also hypothesised that, in Study 1, these findings would be discriminant to implicit happiness, such that there would not be an association between FHS scores and implicit sadness. Indeed, we did not find a significant relationship between self-reported fear of happiness and implicit sadness in the multilevel models, providing evidence for the discriminant validity of these happiness attributions.

We implemented a mood induction procedure (MIP) in Studies 2 and 3 and hypothesised that inducing a positive mood in participants, in general, would result in greater implicit happiness. When investigating this hypothesis in Study 2, we did not find an effect of the positive mood induction on self-reported fear of happiness or implicit happiness. Our results from Study 3, however, did support this hypothesis: individuals in the positive mood induction endorsed an increase in happiness from pre- to post-induction, and this finding was discriminant from other emotions (e.g. anger, fear, and sadness). In addition, individuals in the neutral mood induction did not endorse any change in emotions from pre- to post-induction.

We next hypothesised that implicit happiness would be endorsed *less* after the positive mood induction for those who fear happiness, extending from our hypothesis above that reduced implicit happiness

Table 5. Intercorrelations of Average IMDES Variables and FHS in Study 3.

Measure	1.	2.	3.	4.	5.
1. FHS	-				
2. Anger	.11**	-			
3. Fear	-.01	.09*	-		
4. Sadness	.06	.09*	.08*	-	
5. Happiness	-.07	-.25***	-.15**	-.13**	-

Note. $N = 511$. *$p < .05$, **$p < .01$, ***$p < .001$

Table 6. Hierarchical Multiple Regression Analyses Predicting FHS from IMDES Emotions in Study 3.

Predictor	ΔR^2	β
Step 1	.015	
Anger		.11*
Fear		-.02
Sadness		.06
Step 2	.002	
Anger		.10*
Fear		-.03
Sadness		.05
Happiness		-.05

Note. $N = 511$. *$p < .05$.

would be associated with greater explicit fear of happiness. Overall, participants in both Studies 2 and 3 rated more images as conveying happiness compared to all other emotions (e.g. anger, fear, and sadness). Importantly, however, this relationship was *opposite* for individuals with greater fear of happiness in Study 2. In other words, the motivational tendency to avoid positivity was discriminant to *only* those with a greater fear of happiness. We did not find this same pattern in Study 3, however, which would have replicated Study 2's findings. It is possible that individuals who are fearful of happiness may have had feelings of *anger* evoked from the mood inductions. Specifically, they may have become irritated or annoyed at the thought of participating in a study that would help them increase happiness or maintain their current emotions. Indeed, fear and anger demonstrate similar levels of valence and arousal.

However, whereas individuals overall in Study 3 endorsed greater implicit happiness after receiving the positive mood induction, this finding did not emerge when including fear of happiness in the model. This may suggest that the positive mood induction was not effective for individuals who have negative views about positivity and happiness. Indeed, this is supported by prior research that some participants dampen or devalue positivity when they are not able to avoid experiencing positivity, which may thus be why we did not see changes in emotions from pre- to post-assessment for individuals who fear happiness (Raes et al., 2012; Winer & Salem, 2016).

Lastly, we examined the role of response time in the relationship between implicit happiness and explicit fear of happiness. We hypothesised in Study 1 that this relationship would significantly interact with response time, such that faster respondents on the implicit measure would be more likely to show this association. However, our findings did not support this hypothesis as we did not find a relationship between response time, self-reported fear of happiness, and implicit happiness. Interestingly, the repeated measures correlation between happiness attributions and FHS scores was not significant (see Table 1). However, the correlation between these two variables was significant at each time point (see Tables S1-S4). Repeated measures correlations analyze the common intra-individual association between two measures taken at various measurement occasions, which helps take into account the non-independence of the data (Bakdash & Marusich, 2017). As such, this process is similar to a random intercepts-only multilevel model. In our main multilevel model (discussed further in the supplemental materials), the intraclass correlation coefficient (ICC) suggested that more variance in the outcome (i.e. fear of happiness scores) was explained by between-subject variability (i.e. the fixed effects) as opposed to within-subject variability. This proportion of variation is likely expressed in the repeated measure correlation and may explain the non-significance of this finding. However, the benefit of multilevel models is that they can simultaneously analyze different sources of variance via fixed and random effects, which lends confidence to the discriminant findings reported in Study 1 (i.e. happiness attributions significantly predicting fear of happiness scores while effectively controlling for sadness attributions).

Although the findings from Study 1 did not provide support for automatic emotional processing of the presented stimuli, the updated design of the implicit measure used in Studies 2 and 3, which required participants to respond in the last 3 s (rather than as quickly as possible), may provide support for a more *active* avoidance of reward. This deliberate decision-making process is also in line with RDT, which posits that individuals do not simply lack an ability to value reward; instead, they actively inhibit this reward (Winer & Salem, 2016). Thus, the designs of the implicit measures used in these studies may instead merely tap into one's unconscious motivations to avoid happiness, as participants are more likely to rely on their own emotions to make emotional attributions to the Chinese characters or abstract paintings (Bartoszek & Cervone, 2017, 2022).

Overall, the discriminant findings from these studies provide further evidence for an avoidance of reward that is present for individuals who fear happiness and positivity, as suggested by RDT (Winer & Salem, 2016). Individuals who possibly hold a heuristic that positive information is associated with negativity or disappointing outcomes also evidence a pattern of unconscious avoidance of happiness on an implicit measure of emotions. Moreover, this pattern was not associated with other emotions that may be seen as negative (i.e. sadness), further corroborating this devaluative process. Whereas these findings provide important information of the processes associated with avoiding positivity, they also have important implications for emotion regulation. Individuals who fear happiness may engage in implicit avoidance behaviours to regulate their experiences of negative emotions associated with positivity.

Strengths and limitations

A clear initial strength is the replicability of these findings across three distinct samples, one that assessed these relationships in a sample of Mturk workers over time, and two that assessed these relationships cross-sectionally in student samples. Taken together, these findings provide compound evidence for replicability and generalizability. Another strength of this study is the discriminant validity of the implicit measure given that reduced happiness attributions, and not sadness attributions, were related to explicit fear of happiness. In addition, the time constraint allowed us to capture participants' automatic processing to examine these relationships. Therefore, future research can benefit from further use of the psychometrically sound implicit measure used in these studies.

A potential weakness concerns the sample demographics for the current studies. Specifically, the samples in all three studies were predominantly White and from the United States, and the samples in Study 2 and Study 3 were college students, potentially limiting the generalizability of our findings. Indeed, individuals in Eastern cultures have different views of happiness and could thus respond differently to both the IMDES and FHS (Joshanloo, 2014). In addition, although prior research indicates that MTurk is an effective and reliable method for collecting data regarding clinical and affective variables (Chandler & Shapiro, 2016; Shapiro et al., 2013), it is possible that we were not able to fully identify careless or sham responding with our validity item in Study 1. Relatedly, a large number of participants in our studies were not included in the current analyses due to them not passing the validity check (Studies 1 and 3) failing to complete all time points (Study 1), guessing the purpose of the mood induction (Studies 2 and 3), or not watching the full mood induction (Study 3), which could potentially bias our results. However, as noted above, we examined patterns of missingness (see Supplemental Materials) and found that the data are missing completely at random (e.g. not due to differences in demographics or our variables of interest). In addition, we believe that our stringent criteria for inclusion does far more to strengthen than harm the validity of our findings.

In addition, the means of the FHS in the present samples suggest the samples were not enriched for clinically significant depressive symptoms (e.g. Gilbert et al., 2014). Given that fearing happiness is likely discriminately predictive even among those with other symptoms of depression and anxiety (Winer & Salem, 2016), future research should examine implicit measures in relation to fear of happiness in clinical populations. A related limitation for the IMDES is that this measure only allows for the assessment of one emotion per item at a time (e.g. participants cannot determine if both "anger" and "happiness" correspond to the same stimulus). Indeed, "happiness" and "anger" were related to FHS scores in study 2 and 3, respectively, suggesting the mood induction may potentially elicit either or both emotions in those who fear happiness; however, it is unclear whether the former or the latter is the case with the design of the current studies. Ultimately, examining whether anger and happiness are co-occurring (or occurring very closely together), such that happiness leads to associated anger (due to negative views of happiness), would require an implicit measure allowing for mixed emotional responses.

An important limitation of Study 2 is the lack of effect from the mood induction. As noted above, there may have been a positivity offset present given that participants demonstrated higher levels of implicit happiness, which may have contributed to our null findings. Meta-analytic findings suggest that using a film with instructions is effective for inducing a positive mood (Joseph et al., 2020), thus we expected that our mood induction would be effective. However, our Study 3 replicated methods from Study 2 and demonstrated that the mood induction was effective with a larger sample size. Additionally, we did not observe the same positivity offset in Study 3 as the emotion endorsements before the mood induction were similar to each other, which allowed the endorsement of happiness to increase after the positive mood induction only.

Conclusion

In sum, findings from two separate studies suggest that those who make fewer happiness attributions on an implicit measure of distinct emotions are more likely to endorse greater, explicit fear of happiness, and findings from one study suggest that individuals who endorse explicit fear of happiness have a dampening or avoidant response to positivity. These findings were also notably discriminant to reward (e.g. happiness), providing further evidence that those who hold negative views of positivity process happiness in a manner that is related to devaluing reward. Future research can further examine how implicit behavioural measures assessing devaluation and avoidance of positivity relate to other clinically-relevant person variables (e.g. anhedonia).

Notes

1. Although individuals are required to give a perception or attribution of what emotion is being conveyed in the ambiguous stimulus, we use the word "implicit" in the current study to describe the response that an individual gives in response to an ambiguous stimulus. This is in part due to the automatic, unconscious processing of these tasks such that participants often rely on their own emotions as a source of information.
2. Although the implicit measure in Study 1 used Chinese characters, we did not assess for whether participants were familiar with or spoke Chinese.
3. Given these data come from a convenience sample, coupled with a lack of a clear consensus on the parameters needed to estimate power for multilevel models (Lane & Hennes, 2017), no formal a priori power analysis was conducted.
4. The procedure of Study 2 was part of a larger study examining behavioural indices of avoidance. Thus, we did not include these face-valid questions to reduce potential demand effects on the behavioural tasks that followed the MIP and IMDES.
5. The scores for each of the 4 emotions (anger, fear, sadness, and happiness) can vary between 0 and 20. Thus, the neutrality would be indicated by a value of 0 for a given emotion score. The "none" response makes it possible for all emotion scores to be zero, as respondents can rate all pictures as expressing no emotions. Moreover, because respondents have to choose only one response option for each picture, including all 5 scores in the analyses would be redundant, as the "none" score can be determined once the other four scores are known.
6. FHS scores were not available for two participants due to them having missing data.
7. Given that the four emotions are not exclusive of one another, we also conducted a rmANOVA to investigate the effect of MIP on the endorsement of happiness. We included MIP as the between-subject factor and assessment period as the within-subject factors. Our findings support the results from our mixed ANOVA: there was neither a significant main effect of assessment period on implicit happiness, $F(1, 226) = .39$, $p = .54$, $\eta_p^2 = .002$, nor a significant interaction of assessment period and MIP on implicit happiness, $F(1, 226) = .36$, $p = .55$, $\eta_p^2 = .002$. Including FHS as a covariate did not alter our findings.
8. We initially planned to run this study in person to better replicate Study 2; however, we were unable to do so due to COVID-19 interrupting in-person data collections. Thus, we ran the study online via Qualtrics.
9. Given that the four emotions are not exclusive of one another, we also conducted a rmANOVA to investigate the effect of MIP on the endorsement of happiness. We included MIP as the between-subject factor and assessment period as the within-subject factors. Our findings support the results from our mixed ANOVA: there was a main effect of assessment period on implicit happiness, $F(1, 509) = 7.32$, $p = .007$, $\eta_p^2 = .01$, and a significant interaction of assessment period and MIP on implicit happiness, $F(1, 509) = 20.44$, $p < .001$, $\eta_p^2 = .04$. Including FHS as a covariate revealed a significant interaction of assessment period on MIP, $F(1, 507) = 5.02$, $p = .03$, $\eta_p^2 = .01$, but all other main effects and interactions were not significant.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

The author(s) reported there is no funding associated with the work featured in this article.

ORCID

Amanda C. Collins http://orcid.org/0000-0002-8258-2272
D. Gage Jordan http://orcid.org/0000-0002-4374-016X
Gregory Bartoszek https://orcid.org/0000-0003-0814-5100
Alisson N. S. Lass https://orcid.org/0000-0002-5263-7228
E. Samuel Winer https://orcid.org/0000-0002-3817-874X

References

Arieti, S., & Bemporad, J. (1980). *Severe and mild depression: The psychotherapeutic approach*. Tavistock.

Bakdash, J. Z., & Marusich, L. R. (2017). Repeated measures correlation. *Frontiers in Psychology*, *8*, 456. https://doi.org/10.3389/fpsyg.2017.00456

Bartoszek, G. (2017). *Development and validation of the implicit measure of distinct emotional states (IMDES).* (Doctoral dissertation). University of Illinois at Chicago, Illinois, Chicago.

Bartoszek, G., & Cervone, D. (2017). Toward an implicit measure of emotions: Ratings of abstract images reveal distinct emotional states. *Cognition and Emotion*, *31*(7), 1377–1391. https://doi.org/10.1080/02699931.2016.1225004

Bartoszek, G., & Cervone, D. (2022). Measuring distinct emotional states implicitly: The role of response speed. *Emotion*, *22*(5), 954–970. https://doi.org/10.1037/emo0000894

Bates, D., Kliegl, R., Vasishth, S., & Baayen, H. (2015). *Parsimonious mixed models*. https://arxiv.org/abs/1506.04967.

Brownstein, M., Madva, A., & Gawronski, B. (2020). Understanding implicit bias: Putting the criticism into perspective. *Pacific Philosophical Quarterly*, *101*(2), 276–307. doi:10.1111/papq.12302

Bryant, J., Winer, E. S., Salem, T., & Nadorff, M. R. (2017). Struggling toward reward: Recent experience of anhedonia interacts with motivation to predict reward pursuit in the face of a stressful manipulation. *PLOS ONE*, *12*(3), e0173439. https://doi.org/10.1371/journal.pone.0173439

Cacioppo, J. T., & Berntson, G. G. (1994). Relationship between attitudes and evaluative space: A critical review, with emphasis on the separability of positive and negative substrates. *Psychological Bulletin*, *115*(3), 401–423. https://doi.org/10.1037/0033-2909.115.3.401

Cham, H., Reshetnyak, E., Rosenfeld, B., & Breitbart, W. (2017). Full information maximum likelihood estimation for latent variable interactions with incomplete indicators. *Multivariate Behavioral Research*, *52*(1), 12–30. https://doi.org/10.1080/00273171.2016.1245600

Chandler, J., & Shapiro, D. N. (2016). Conducting clinical research using crowdsourced convenience samples. *Annual Review of Clinical Psychology*, *12*(1), 53–81. https://doi.org/10.1146/annurev-clinpsy-021815-093623

De Houwer, J., & Moors, A. (2010). Implicit measures: Similarities and differences. In B. Gawronski, & K. B. Payne (Eds.), *Handbook of implicit social cognition: Measurement, theory, and applications* (pp. 176–193). Guilford Press.

De Houwer, J., Teige-Mocigemba, S., Spruyt, A., & Moors, A. (2009). Implicit measures: A normative analysis and review. *Psychological Bulletin*, *135*(3), 347–368. https://doi.org/10.1037/a0014211

Evans, J. S. B., & Stanovich, K. E. (2013). Dual-process theories of higher cognition: Advancing the debate. *Perspectives on Psychological Science*, *8*(3), 223–241. https://doi.org/10.1177/1745691612460685

Feldman, G. C., Joormann, J., & Johnson, S. L. (2008). Responses to positive affect: A self report measure of rumination and dampening. *Cognitive Therapy and Research*, *32*(4), 507–525. https://doi.org/10.1007/s10608-006-9083-0

Finch, W. H., Bolin, J. E., & Kelley, K. (2016). *Multilevel modeling using R*. CRC Press.

Forgas, J. P. (1995). Mood and judgment: The affect infusion model (AIM). *Psychological Bulletin*, *117*(1), 39–66. https://doi.org/10.1037/0033-2909.117.1.39

Gawronski, B. (2009). Ten frequently asked questions about implicit measures and their frequently supposed, but not entirely correct answers. *Canadian Psychology*, *50*(3), 141–150. https://doi.org/10.1037/a0013848

Gawronski, B., & De Houwer, J. (2014). Implicit measures in social and personality psychology. In H. T. Reis, & C. M. Judd (Eds.), *Handbook of research methods in social and personality psychology* (Vol. 2, pp. 283–310). Cambridge University Press. https://doi.org/10.1017/CBO9780511996481.

Gawronski, B., & Hahn, A. (2019). Implicit measures: Procedures, use, and interpretation. In H. Blanton, J. M. LaCroix, & G. D. Webster (Eds.), *Measurement in social psychology* (pp. 29–55). Routledge/Taylor & Francis Group.

Gelman, A., & Hill, J. (2006). *Data analysis using regression and multilevel/hierarchical models*. Cambridge University Press.

Gilbert, P., McEwan, K., Catarino, F., & Baião, R. (2014). Fears of negative emotions in relation to fears of happiness, compassion,alexithymia and psychopathology in a depressed population: A preliminary study. *Journal of Depression and Anxiety*, *SS2*(004), https://doi.org/10.4172/2167-1044.S2-004

Gilbert, P., McEwan, K., Catarino, F., Baião, R., & Palmeira, L. (2014). Fears of happiness and compassion in relationship with depression, alexithymia, and attachment security in a depressed sample. *British Journal of Clinical Psychology*, *53*(2), 228–244. https://doi.org/10.1111/bjc.12037

Gilbert, P., McEwan, K., Gibbons, L., Chotai, S., Duarte, J., & Matos, M. (2012). Fears of compassion and happiness in relation to alexithymia, mindfulness, and self-criticism. *Psychology and Psychotherapy: Theory, Research and Practice*, *85*(4), 374–390. https://doi.org/10.1111/j.2044-8341.2011.02046.x

Greifeneder, R., Bless, H., & Pham, M. T. (2011). When do people rely on affective and cognitive feelings in judgment? A review. *Personality and Social Psychology Review*, *15*(2), 107–141. https://doi.org/10.1177/1088868310367640

Harrison, E. (2022). *Package "finalfit"*. Retrieved November 29, 2022 from https://cran.r-project.org/web/packages/finalfit/finalfit.pdf.

Heck, R. H., Thomas, S. L., & Tabata, L. N. (2013). *Multilevel and longitudinal modeling with IBM SPSS*. Routledge.

Hox, J. J., Moerbeek, M., & van de Schoot, R. (2018). *Multilevel analysis: Techniques and applications* (3rd ed.). Routledge.

Jamshidian, M., & Jalal, S. (2010). Tests of homoscedasticity, normality, and missing completely at random for incomplete multivariate data. *Psychometrika*, *75*(4), 649–674. https://doi.org/10.1007/s11336-010-9175-3

Jamshidian, M., Jalal, S., & Jansen, C. (2014). MissMech: An R package for testing homoscedasticity, multivariate normality, and missing completely at random (MCAR). *Journal of Statistical Software*, *56*(6), 1–31. https://doi.org/10.18637/jss.v056.i06

Jordan, D. G., Winer, E. S., Salem, T., & Kilgore, J. (2018). Longitudinal evaluation of anhedonia as a mediator of fear of positive evaluation and other depressive symptoms. *Cognition and Emotion*, *32*(7), 1437–1447. https://doi.org/10.1080/02699931.2017.1289895

Joseph, D. L., Chan, M. Y., Heintzelman, S. J., Tay, L., Diener, E., & Scotney, V. S. (2020). The manipulation of affect: A meta-analysis of affect induction procedures. *Psychological Bulletin*, *146*(4), 355–375. https://doi.org/10.1037/bul0000224

Joshanloo, M. (2014). Eastern conceptualizations of happiness: Fundamental differences with Western views. *Journal of Happiness Studies*, *15*(2), 475–493. https://doi.org/10.1007/s10902-013-9431-1

Joshanloo, M., Lepshokova, Z. K., Panyusheva, T., Natalia, A., Poon, W.-C., Yeung, V. W., Sundaram, S., Achoui, M., Asano, R., Igarashi, T., Tsukamoto, S., Rizwan, M., Khilji, I. A., Ferreira, M. C., Pang, J. S., Ho, L. S., Han, G., Bae, J., & Jiang, D.-Y. (2013). Cross-cultural validation of Fear of Happiness Scale across 14 national groups. *Journal of Cross-Cultural Psychology*, *45*(2), 246–264. https://doi.org/10.1177/0022022113505357

Joshanloo, M., & Weijers, D. (2014). Aversion to happiness across cultures: A review of where and why people are averse to happiness. *Journal of Happiness Studies*, *15*(3), 717–735. https://doi.org/10.1007/s10902-013-9489-9

Kline, R. B. (2015). *Principles and practice of structural equation modeling* (4th ed.). Guilford Press.

Lane, S. P., & Hennes, E. P. (2017). Power struggles. *Journal of Social and Personal Relationships*, *35*(1), 7–31. doi:10.1177/0265407517710342

Leeper, T. J. (2014). *Introduction to the simple wizard (text based)*. https://github.com/leeper/MTurkR/wiki/Wizard-Text-Based.

Lo, S., & Andrews, S. (2015). To transform or not to transform: Using generalized linear mixed models to analyse reaction time data. *Frontiers in Psychology*, *6*, 1171–1187. https://doi.org/10.3389/fpsyg.2015.01171

Mauss, I. B., & Robinson, M. D. (2009). Measures of emotion: A review. *Cognition & Emotion*, *23*(2), 209–237. https://doi.org/10.1080/02699930802204677

Murphy, J., Catmur, C., & Bird, G. (2018). Alexithymia is associated with a multidomain multidimensional failure of interoception: Evidence from novel tests. *Journal of Experimental Psychology: General*, *147*(3), 398–408. doi:10.1037/xge0000366

Palkovics, M. A., & Takáč, M. (2016). Exploration of cognition–affect and Type 1–Type 2 dichotomies in a computational model of decision making. *Cognitive Systems Research*, *40*, 144–160. https://doi.org/10.1016/j.cogsys.2016.06.001

Pinheiro, J., & Bates, D. (2000). *Mixed-effects models in S and S-PLUS*. Springer.

Pinheiro, J., Bates, D., DebRoy, S., & Sarkar, D. (2018). *nlme: Linear and nonlinear mixed effects models. R package version 3.1-137*. https://cran.r-project.org/package=nlme.

Pool, E., Brosch, T., Delplanque, S., & Sander, D. (2016). Attentional bias for positive emotional stimuli: A meta-analytic investigation. *Psychological Bulletin*, *142*(1), 79–106. https://doi.org/10.1037/bul0000026

Quoidbach, J., Berry, E. V., Hansenne, M., & Mikolajczak, M. (2010). Positive emotion regulation and well-being: Comparing the impact of eight savoring and dampening strategies. *Personality and Individual Differences*, *49*(5), 368–373. https://doi.org/10.1016/j.paid.2010.03.048

Raes, F., Smets, J., Nelis, S., & Schoofs, H. (2012). Dampening of positive affect prospectively predicts depressive symptoms in non-clinical samples. *Cognition & Emotion*, *26*(1), 75–82. https://doi.org/10.1080/02699931.2011.555474

Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology*, *39*(6), 1161–1178. https://doi.org/10.1037/h0077714

Russell, J. A. (2003). Core affect and the psychological construction of emotion. *Psychological Review*, *110*(1), 145–172. doi:10.1037/0033-295X.110.1.145

Şar, V., Türk, T., & Öztürk, E. (2019). Fear of happiness among college students: The role of gender, childhood psychological trauma, and dissociation. *Indian Journal of Psychiatry*, *61*(4), 389–394. doi:10.4103/psychiatry

Schwarz, N. (2011). Feelings-as-information theory. In P. Van Lange, A. Kruglanski, & E. T. Higgins (Eds.), *Handbook of theories of social psychology: Volume one* (pp. 289–308). Sage Press.

Shapiro, D. N., Chandler, J., & Mueller, P. A. (2013). Using Mechanical Turk to study clinical populations. *Clinical Psychological Science*, *1*(2), 213–220. https://doi.org/10.1177/2167702612469015

Siemer, M., & Reisenzein, R. (1998). Effects of mood on evaluative judgements: Influence of reduced processing capacity and mood salience. *Cognition & Emotion*, *12*(6), 783–805. https://doi.org/10.1080/026999398379439

Tamir, M., Ford, B. Q., & Ryan, E. (2013). Nonconscious goals can shape what people want to feel. *Journal of Experimental Social Psychology*, *49*(2), 292–297. https://doi.org/https://doi.org/10.1016/j.jesp.2012.11.009

Tierney, N., Di, C., McBain, M., & Fay, C. (2021). *naniar: Data structures, summaries, and visualisations for missing data*. Retrieved November 29, 2022 from https://cran.r-project.org/web/packages/naniar/index.html.

Winer, E. S., & Salem, T. (2016). Reward devaluation: Dot-probe meta-analytic evidence of avoidance of positive information in depressed persons. *Psychological Bulletin*, *142*(1), 18–78. https://doi.org/10.1037/bul0000022