The complex misestimation of others' emotions: Underestimation of emotional prevalence versus overestimation of emotional intensity and their associations with well-being

Ho Ching Ip
ORCID: 0000-0002-9549-7544
Department of Psychology, University of Hong Kong
u3569377@connect.hku.hk / melodyhcip@gmail.com

^Gilad Feldman
OCRID: 0000-0003-2812-6599
Department of Psychology, University of Hong Kong, Hong Kong SAR
gfeldman@hku.hk / giladfel@gmail.com

^Corresponding author

Author bios:

Ho Ching Ip was a student at the University of Hong Kong during the academic year 2022-2023.

Gilad Feldman is an assistant professor at the University of Hong Kong psychology department. His research focuses on judgment and decision-making.

Declaration of Conflict of Interest:

Financial disclosure/funding:

Authorship declaration:

Ho Ching Ip conducted the replication as part of her thesis.

Gilad Feldman was the thesis supervisor and led the replication efforts. He supervised each step in the project, conducted the pre-registrations, ran data collection, revised and edited the manuscript for final submission.

Corresponding author:

Gilad Feldman, Department of Psychology, University of Hong Kong, Hong Kong SAR; gfeldman@hku.hk ; 0000-0003-2812-6599

Rights:

Acknowledgements:

Contributor Roles Taxonomy

Role	Ho Ching Ip	Gilad Feldman
Conceptualization	X	X
Pre-registration	X	X
Data curation		X
Formal analysis	X	
Funding acquisition		X
Investigation	X	
Pre-registration peer review / verification		X
Data analysis peer review / verification		X
Methodology	X	
Project administration		X
Resources		X
Software	X	
Supervision		X
Validation		X
Visualization	X	
Writing-original draft	X	
Writing-review and editing		X

Abstract

Jordan et al. (2011) demonstrated that people underestimated the prevalence of others' negative emotional experiences and that these were associated with higher well-being. We conducted a pre-registered replication of Studies 1b and 3 by Jordan et al. (2011) ($N = 594$) with adjustments and added extensions. Building on their methodology we examined both prevalence and intensity of emotional experiences, and our findings suggest a much more complex story with surprising effects. We found an underestimation of the prevalence of negative emotions, but also unexpectedly of an underestimation of the prevalence of positive emotions, with stronger effects for negative than for positive emotions. However, we found an opposite effect for emotional intensity, people overestimated the intensity of both positive and negative emotional experiences, again with stronger effects for negative. Surprisingly, associations between prevalence estimations and well-being were in the opposite direction to the target article's. Materials, data, and code: https://osf.io/bwmtr/

The complex misestimation of others' emotions:

Underestimation of emotional prevalence versus overestimation of emotional intensity and

their associations with well-being

Background

A growing body of literature has documented people's misperceptions of others' attitudes, beliefs, and behaviors (Prentice & Miller, 1993). These errors in perceiving others' psyche and mental states seem at least in part related to purposeful public misrepresentations, as people misrepresent their private worlds to meet what they perceive social norms dictate. For example, people tend to underestimate their peers' negative emotional experiences, as people tend to hide their negative emotions (Larson et al., 1982) to adhere to perceived display rules to appear positive (Ekman & Friesen, 1969).

Jordan et al. (2011) demonstrated that people underestimated the prevalence of peers' negative emotional experiences more than their peers' positive emotions[1], and that prevalence estimates were associated with a host of well-being measures. Their findings showed that those with lower prevalence estimates of negative emotions reported greater loneliness, greater rumination, and less satisfaction with life.

We aimed to revisit and extend Jordan et al. (2011)'s theory and findings. Our first goal was to conduct an independent well-powered pre-registered close replication examining misestimations of emotional prevalence and their associations with psychological well-being,

---

[1] This is our reframing of the target's original framing claiming no misestimation of positive emotions, a null hypothesis, more on that below.

adopting and improving on the target article's methodology. Our second goal was to add extensions and examine: 1) whether misestimations extend beyond prevalence to intensity, and 2) whether social comparison orientation moderates associations between prevalence and intensity and well-being.

We begin by introducing the literature on emotional misestimations and the chosen article for replication - Jordan et al. (2011). We then outline our chosen studies for replication from the target article, the target's experimental designs, and our adaptations, improvements, and extensions.

The misestimations of others' emotions

Misestimations of others' views, behaviors, and experiences were initially mostly studied in the context of bystander nonintervention during emergencies (Allport, 1924; Darley & Latane, 1968). It was later developed to study people's misinterpretations of social norms based on their observations of public behaviors, which may be misaligned with people true private feelings, especially when those are considered negative (Prentice & Miller, 1993).

One of the earliest demonstrations of the differences between public behaviors and private attitudes was in the 1930s, showing a misalignment between what people said in public compared to what they reported in private settings (Schanck, 1932). In a review by Sargent and Newman (2021), they summarized three areas in which this phenomenon has been demonstrated - drug use, alcohol, and sexual and dating norms. For instance, participants in the study of Prentice and Miller (1993) expressed in private that they were uncomfortable with excessive drinking, yet publicly endorsed pro-drinking policies, presumably due to misperceiving their peers' behaviors and attitudes and being pro-drinking.

Jordan et al. (2011) extended the literature by demonstrating people's misestimations of the prevalence of others' negative emotions. The misestimation is likely due to people's tendency to try to suppress their negative emotions in social contexts (Larson et al., 1982), given that they are deemed less socially appropriate and not in line with norms for social behavior (Ekman & Friesen, 1969).

Choice of study for replication: Jordan et al. (2011)

We aimed to revisit the classic phenomenon to examine the reproducibility and replicability of the findings with an independent, well-powered, pre-registered close replication of Jordan et al. (2011) with several extensions. We answered recent calls and a growing recognition of the importance of reproducibility and replicability in psychological science (e.g., Nosek et al., 2022; Zwaan et al., 2018).

We chose the article by Jordan et al. (2011) as the target for replication based on several factors: 1) its academic and practical impact with no direct replications, 2) the potential for improvement on their methods and analyses, 3) the potential in answering new theoretical questions using an extension in examining misestimation of emotional intensity.

The article has had an impact on scholarly research in the area of social psychology, with 306 Google Scholar citations at the time of writing (August 2024). The article was followed by empirical research demonstrating mis-estimations and their impact in social media settings (Tandoc et al., 2015), job-seeking (Burke & Kraut, 2013), risky sexual behaviors (Young & Jordan, 2013), and impulsive buying (Liu et al., 2019). Other research extended the original findings and further investigated the practical implications of the associations between mis-estimations and psychological well-being outcomes, such as for depressive symptoms (Steers et

al., 2014) and self-esteem (Alfasi, 2019; Wang et al., 2017). To the best of our knowledge, there are currently no published direct independent replications of the chosen target article.

When we analyzed the article, we realized that some analyses were in need of revision and improvement. The analyses conducted for these studies were on an item-level, with six items for positive, and six items for negative, making it very challenging to detect any effects, which is related to what seemed like unreasonably large effects ($d = 2.52$ in Study 1b, and $d = 3.16$ in Study 3). It was not clear to us why the analyses were run only on an item-level rather than also on a participant-level. Given the target article's null hypothesis for positive emotions, we were open to the possibility of also finding support for meaningful effects for positive emotions and that with a larger sample of items or analyses on a participant-level, the difference from the null might be more easily detected. Given our surprising findings we report below, we also sought an external expert reviewer to examine the regression analyses conducted in Study 3 of the target article (shared on https://osf.io/zy5qa/). In his analyses, the external reviewer raised the possibility of a suppression effect in the regression analyses between prevalence estimates and well-being (Sharpe & Roberts, 1997; Thompson & Levine, 1997). We, therefore, saw the potential in improving on the conducted analyses by also conducting participant-level analyses and focusing on raw correlations.

We also sought to go beyond examining the misestimation of emotional prevalence to also examine an important missing piece of the puzzle – the possible misestimation of emotional intensity. We built on the same methodology to examine both aspects of emotions together, resulting in a more comprehensive view of a much broader phenomenon.

Original hypotheses and findings in target article: Jordan et al. (2011)

Jordan et al. (2011) empirical work consisted of four studies, and in the current replication we focused on Studies 1b and 3.

In Study 1b, they examined whether people would accurately predict the prevalence of others' emotional experiences. Participants rated their own experiences and estimated the prevalence of six positive emotional experiences (such as receiving high grades or attending a fun party) and six negative emotional experiences (such as thinking about workload or having a fight/argument) in the two weeks prior. The researchers hypothesized that (a) people underestimate negative experiences, yet that (b) these errors are not present for positive experiences. We note that their second hypothesis regarding positive emotions was a null hypothesis, which we believe is better reframed to an interaction hypothesis: (b-reframed) underestimation errors are stronger for negative experiences than for positive experiences. Study 2 showed similar results using a re-analysis of a dataset by Srivastava et al. (2009) with an underestimation of negative emotions and an overestimation of positive emotions.

In Study 3, the researchers examined the associations between prevalence estimations and psychological well-being. Participants completed scales regarding loneliness, rumination, depressive symptoms, life satisfaction, subjective happiness, and self-reported their number of friends, in addition to rating the same emotional events from Study 1b. They hypothesized that (c) higher prevalence estimates of others' negative emotional experiences are associated with poorer psychological outcomes: greater loneliness, rumination, depressive symptoms, and lower life satisfaction and subjective happiness. They also hypothesized that (d) having fewer peers is associated with stronger mis-estimation.

Table 1

Jordan et al. (2011) Studies 1b and 3: Summary of findings

| Experiences | Study 1b | | | Study 3 | | |
	Estimation Error[1]	Average Estimation Error	t-statistics	Estimation Error	Average Estimation Error	t-statistics
Negative experiences		-17.2%	5.47**		-21.4%	5.99**
Had fight/argument	-13.0***			-13.8***		
Thought about distant friends/ family	-28.2***			-26.3***		
Thought about enormous workload	-12.2***			-11.3***		
Was rejected by boy/girl	-8.9***			-18.4***		
Received low grade	-15.9***			-23.3***		
Thought about bad personal health habits	-24.1***			-35.0***		
Positive experiences		+5.6%	1.18		+3.8%	1.06
Received high grade	-3.0			-0.3		
Attended fun party	+20.9***			+13.2***		
Participated in athletics	+13.7***			+7.6***		
Went out with friends	+12.6***			+11.5***		
Talked to distant friends/ family	-8.3***			-9.9***		
Had great meal	-2.3			+0.5		

Note. [1]A positive number indicates an overestimation and a negative number indicates an underestimation. The table was adopted from Jordan et al. (2011), pp. 126, 130. **$p < .01$. ***$p < .001$.

We summarized the findings of Study 1b and Study 3 in Table 1 and the hypotheses of the target article in Table 2 (under "Replication: Prevalence estimations").

Extensions

Intensity of emotional experiences

The authors of the target article acknowledged a limitation in their design, specifically that they did not differentiate between prevalence and intensity. For instance, participants were asked to rate the prevalence of their peers "receiving a bad grade and felt really bad about it", making it difficult to distinguish between their estimation of the prevalence of receiving a bad grade and their estimation of the intensity of the emotions associated with that event. As such, we were unsure whether participants underestimated the prevalence of their peers' negative emotions because they believed their peers rarely experienced such events (prevalence), or because they believed they these events would not elicit strong emotional responses (intensity).

Therefore, we revised the questions to inquire about both the prevalence of the events and the intensity of the emotional experiences. By expanding upon the original article, our aim was to examine the interplay between prevalence and intensity estimates and gain a clearer understanding of the source of the misestimation.

Social Comparison Orientation (Exploratory)

We aimed to investigate the role of social comparison orientation, the tendency to compare oneself with others, as a potential predictor of misestimation and well-being measures. Gibbons and Buunk (1999) found that individuals with a higher social comparison orientation exhibited greater uncertainty and instability regarding their self-worth. Consequently, they tended to evaluate themselves based on how others were faring in public. This has the potential to be even

more pronounced in an era in which social media provides ample opportunities for social comparison.

Accumulating evidence suggests a positive link between social comparison orientation and threats to psychological well-being online, as many social media users tend to selectively present mostly positive aspects of their lives (Vogel et al., 2014). For example, Vogel et al. (2015) found that people with higher social comparison orientation were more susceptible to deleterious consequences such as lower self-esteem and poorer affect balance when using social media. Similarly, Wang et al. (2017) found that social comparison orientation was associated with increased usage of social networks, which in turn predicted lower life satisfaction and higher levels of jealousy and envy.

We therefore considered the possibility that this trait encompasses aspects related to social misestimations, as a greater inclination to compare oneself against others may also make individuals more vulnerable to potential discrepancies between their own experiences and those of others. Therefore, our aim was to examine the association between social comparison orientation, misestimations of others' emotional experiences, and psychological well-being. Furthermore, we sought to explore the interaction between social comparison orientation and misestimation of emotional experiences in relation to well-being measures, with the exploratory prediction that the higher the social comparison orientation, the stronger the negative link between misestimation of emotional experiences with well-being measures.

Replication and Extension: Overview

We summarized the hypotheses for our adjusted replication and extensions in Table 2. In hypotheses 2c and 2d, we built on the target article's hypotheses 2a and 2b to separately assess the prevalence and emotional intensity of emotional events. We also introduced social

comparison orientation as a new measure to examine its association with the misperception of

peers' emotional experiences and psychological well-being in Hypotheses 5 and 6.

Table 2

Replication and extensions: Summary of hypotheses

Replication: Prevalence estimations

#	Hypothesis
2a	People underestimate the prevalence of others' negative emotional experiences.
2b	People do not underestimate the prevalence and extent of others' positive emotional experiences. [Our reframing of the target's null hypothesis: Prevalence underestimation errors are stronger for negative experiences than for positive experiences.]
4	4a) There is a positive association between the estimation of the prevalence of negative emotional experiences and well-being. [1] 4b) There is a negative association between the estimation of the prevalence of positive emotional experiences and well-being. [1] Negative indicators - higher well-being: lower depressive symptoms, lower loneliness, lower rumination
4-1	4-1a) There is a negative association between the estimation of the prevalence of negative emotional experiences and depressive symptoms. 4-1b) There is a positive association between the estimation of the prevalence of positive emotional experiences and depressive symptoms.
4-2	4-2a) There is a negative association between the estimation of the prevalence of negative emotional experiences and loneliness. 4-2b) There is a positive association between the estimation of the prevalence of positive emotional experiences and loneliness.
4-3	4-3a) There is a negative association between the estimation of the prevalence of negative emotional experiences and rumination. 4-3b) There is a positive association between the estimation of the prevalence of positive emotional experiences and rumination. Positive indicators - higher well-being: higher life-satisfaction, higher subjective happiness.
4-4	4-4a) There is a positive association between the estimation of the prevalence of negative emotional experiences and life satisfaction. 4-4b) There is a negative association between the estimation of the prevalence of negative emotional experiences and life satisfaction.
4-5	4-5a) There is a positive association between the estimation of the prevalence of negative emotional experiences and subjective happiness. [2] 4-5b) There is a negative association between the estimation of the prevalence of negative emotional experiences and subjective happiness. [2]

#	Hypothesis
	Extensions: <u>Intensity estimations</u>
2c	People <u>underestimate</u> the <u>intensity</u> of others' <u>negative</u> emotional experiences.
2d	People <u>overestimate</u> the <u>intensity</u> of others' <u>positive</u> emotional experiences.
5	Social comparison orientation interacts with misestimation in predicting well-being: The higher the social comparison orientation, the stronger the associations in Hypothesis 4 (Hypotheses 4-1a to 4-5b).
6	Social comparison orientation is negatively associated with well-being. <u>Negative indicators</u> - higher well-being: lower depressive symptoms, lower loneliness, lower rumination
6a	Social comparison orientation is positively associated with <u>depressive symptoms.</u>
6b	Social comparison orientation is positively associated with <u>loneliness</u>.
6c	Social comparison orientation is positively associated with <u>rumination</u>. <u>Positive indicators</u> - higher well-being: higher life-satisfaction, higher subjective happiness.
6d	Social comparison orientation is negatively associated with <u>life satisfaction</u>.
6e	Social comparison orientation is negatively associated with <u>subjective happiness</u>.

Note. [1] The target article sometimes shifts from referring to misestimations in its predictions to tests regarding prevalence estimations. We aligned the hypotheses with the tests conducted and the findings reported, rather than the versions of theory and hypothesis referring to misestimations. In addition, we specified separate hypotheses for estimations of negative emotions and estimations of positive emotions, to make things clearer.
[2] The hypothesis was not supported in the original target article, yet is included due the possibility of lacking power given the item-level analysis and small number of items.
[3] We had no specific predictions for the associations between intensity estimations of positive and negative emotional experience, and these should therefore be treated as exploratory.

Data availability, pre-registration, and open-science disclosures

We provided all materials, data, and code on the OSF: https://osf.io/bwmtr/. We first pre-registered the experiment on the Open Science Framework (OSF; https://osf.io/qda7j/) and data collection was launched later that week. All measures, manipulations, exclusions are reported, and data collection was completed before analyses. The pre-registration and manuscript were written based on a template by Feldman (2023).

Method

Power analysis and Sensitivity Test

We first calculated the effect sizes of the findings reported in the target article with the help of a guide by Jané et al. (2024). We then conducted an a-priori power analysis (power = 0.95, alpha = 0.05), with an upward adjustment. The calculated effect sizes are summarized in the supplementary materials.

There were several issues with using the target article's effects as a basis for power analyses. The target conducted their analyses on an item-level with an analysis on 12 items, which does not seem to matter for recruit of participants, and with the small number of items requiring a very large effect size on the item-level to be able to detect any effects. This is related to other issues with some of the effects reported in the target (e.g., $d = 2.52$ in Study 1b and $d = 3.16$ in Study 3) being much larger than typical effects in the field (see Jané et al., 2024).

Therefore, instead of using the target article's effects, we aimed for a sample size that would allow for participant-level detection of paired-sample t-test $d(z) = 0.2$ ($N = 272$), and correlations of $r = .15$ ($N = 472$). These effect sizes are far smaller than any of the effects reported in the target article and are considered weak to medium effects in social psychology

(Jané et al., 2024). This approach also aligns with the conservative estimate of 460 according to the rule of thumb recommendations by Simonsohn (2015) for a sample size 2.5 times larger than the combined samples in the original article (80 in the original Study 1b and 104 in the original Study 3), though we note that these recommendations were intended for participant-level between subject-designs, and not for item-level analyses with fixed number of items. To account for possible exclusions in case of a failed replication (see data analysis strategy section) and multiple analyses on the same sample, we exceeded the planned sample by aiming for a sample of 600.

Participants

We recruited a total of 594 US American college students on Prolific (Palan & Schitter, 2018; 290 males, 285 females, 19 other/did not disclose; using the Prolific internal qualifier for students). We used Prolific's filters, by restricting the location to the US using "standard sample", we set it to "Nationality: United States", "Country of birth: United States", "Student status: Yes", "Minimum Approval Rate: 95, Maximum Approval Rate: 100", "Minimum Submissions: 50, Maximum Submissions: 100,000", "Total times a participant can complete your study: Once". We summarized a comparison of the target article sample and the replication sample in Table 3. We outlined the details of the pay and survey procedures in the supplementary.

Table 3

Comparison of target article versus replication: Differences and similarities

	Jordan et al., (2011) Study 1b	Jordan et al., (2011) Study 3	Replication
Sample size	80	104	594
Geographic origin	US American students	US American students	US American students on Prolific
Gender	35 males, 45 females, 0 other/did not disclose	51 males, 54 females, 0 other/did not disclose	290 males, 285 females, 19 other/did not disclose
Medium (location)	Paper-and-pen	Computer (online)	Computer (online)
Compensation	N/A	Nominal payment	Nominal payment (2 British pounds for an estimated 12 minutes completion time)
Year	2011 or earlier	2011 or earlier	2023

Table 4

Replication and extension: Experimental design and measures

IV1: Positive versus negative emotions (within; all participants rated both) IV2: Self versus others rating (within; all participants rated both)	IV1: Estimation of positive emotional events	IV1: Estimation of negative emotional events
	"Felt happy because they…" 1. Received high grades 2. Attended fun party 3. Participated in athletics 4. They went out with friends 5. They talked to distant friends or family 6. Had great meal	"Felt sad because they…" 1. Had a fight or argument 2. Thought about distant friends or family 3. Thought about enormous workload 4. Were rejected by someone 5. Received a low grade 6. Thought about bad personal health habits

Self ratings Participants were asked to estimate based on their own emotional experiences.	DV1: Prevalence and intensity of own emotional experiences [Replication + Extension] "For each of the following emotional experiences, please indicate whether you have experienced those sometime in the past 2 weeks and if you have - the intensity of the emotion. 0 means: You have NOT experienced this emotion in the past two weeks. 1-100 means: You have experienced this emotion at least once in the past two weeks" (0 = Not experienced; 1 = Lowest emotional intensity; 100 = Highest emotional intensity.)
Others ratings Participants were asked to make estimations of US American student participants on Prolific taking the survey.	DV1a: Prevalence of others' emotional experiences [Replication] "Please estimate the percentage of other US American student participants on Prolific taking the survey like you who had had, sometime in the past 2 weeks, each of the following emotional experiences. (0% to 100%) DV1b: Intensity of others' emotional experiences [Extension] "Please try and estimate the emotional intensity for other US American student participants on Prolific taking the survey like you who have experienced this emotion (1 = Lowest emotional intensity; 100 = Highest emotional intensity)"
Exploratory open-ended	Perception of vague test items: Participants briefly wrote about their understanding of vague test items. "What is a "bad" grade? How do you classify a grade as being a bad grade? Can you give a quick example? (1-2 sentences)" "What is an "enormous" workload? How do you classify workload as being an enormous workload? Can you give a quick example? (1-2 sentences)" "What is a "bad" personal health habit? How do you classify personal health habit as being bad? Can you give a quick example? (1-2 sentences)"

Well-being measures and traits	(presented in random order)
	Loneliness [Replication] UCLA Loneliness Scale (Hays & DiMatteo, 1987; 4), 8-item short-form version. Sample items: "There is no one I can turn to." and "I can find companionship when I want it." (1 = *Never* to 4 = *Often*)
	Rumination [Replication] Brooding subscale (5-item) of the Ruminative Responses to Depression Questionnaire (Nolen-Hoeksema, 1991; Treynor et al., 2003). Sample items: "Think "What am I doing to deserve this?" and "Think about a recent situation, wishing it had gone better." etc. (1 = *Almost never* to 4 = *Almost always*)
	Depressive Symptoms [Replication] Center for Epidemiologic Studies Depression Scale 10-item short-form version (Cole et al., 2004; Radloff, 1977). Sample items: "I was bothered by things that usually don't bother me." and "I had trouble keeping my mind on what I was doing." etc. (1 = *Rarely or none of the time (less than 1 day)* to 4 = *All of the time (5-7 days)*)
	Life Satisfaction [Replication] Participants were asked to complete the 5-item Satisfaction with Life Scale (SWLS) (Diener et al., 1985). Sample items: "In most ways my life is close to my ideal." and "If I could live my life over, I would change almost nothing." etc. (1 = *Strongly disagree* to 7 = *Strongly agree*)
	Subjective Happiness [Replication] Subjective Happiness Scale (SHS) 4-item (Lyubomirsky & Lepper, 1999). Sample items: "In general, I consider myself…" and "Compared to most of my peers, I consider myself…" etc. (1 = *Not at all/Less happy*; 7 = *A very happy person/More happy/A great deal*)
	Social Comparison Orientation [Extension] The Iowa-Netherlands Comparison Orientation Measure (INCOM) (Gibbons & Buunk, 1999) Sample items: "I often compare myself with others with respect to what I have accomplished in life" and "If I want to learn more about something I try to find out what others think about it" (1 = *Strongly disagree* to 5 = *Strongly agree*)

Note. All materials are provided in the Qualtrics survey export provided in the OSF folder.

Design and procedure

We summarized the experimental design and all measures in Table 4 (more details are available in the supplementary materials).

Participants rated their own emotional experiences and others' emotional experiences in random order. Misestimations were measured as participants' estimates of others' emotional experiences minus the average of all participants' own emotional experiences. In the target article's Study 1b, the prevalence and intensity were confounded, which we adjusted to measure prevalence and intensity separately (see "Measures" subsection).

We conducted the study using Qualtrics. All participants first indicated their consent. They then rated the prevalence of six positive and six negative emotional experiences for themselves and for other US student participants taking this survey, in counterbalanced order.

Prior to each rating task, we added comprehension checks to ensure that participants were paying attention to the type of rating (prevalence or intensity) and who they are rating (self versus other). Participants had to answer these checks correctly in order to proceed to the rating task. We note that this is a deviation from the target article's procedure to ensure that participants were attentive and knew who and what they were rating.

For exploratory purposes, we included three questions that asked participants to provide short sentences regarding their perceptions of the test items they had previously rated, to help us better understand participants' mindsets when answering some of the items from the target article, to address the possibility of puzzling or surprising findings using these items.

Participants then completed measures assessing their loneliness, rumination tendency, depressive symptoms, life satisfaction, subjective happiness and social comparison orientation.

We randomized the order of the six well-being and trait measures. Finally, participants answered a funneling section, provided demographic information, and were debriefed.

Measures

Prevalence and intensity of own emotional experiences (replication and extension)

In the target article, participants only rated whether they had experienced an emotional event or not. We adjusted and extended the measures to also include intensity:

> "For each of the following emotional experiences, please indicate whether you have experienced those sometime in the past 2 weeks and, if you have, the intensity of the emotion.
>
> 0 means: You have NOT experienced this emotion in the past two weeks.
>
> 1-100 means: You have experienced this emotion at least once in the past two weeks.
>
> 1 = lowest emotional intensity, and 100 = highest emotional intensity."

In this way, we measured both prevalence and intensity in the same question: a value of 0 indicated that the participant has not experienced this event, and a value of 1 to 100 indicated that the participant has experienced it with a rating of the intensity of the experience.

Prevalence of others' emotional experiences (replication)

We closely followed the target's measure of participants' estimates of others' emotional experiences with: "Please estimate the percentage of other US American student participants on Prolific taking the survey like you who had had, sometime in the past 2 weeks, each of the following emotional experiences. (0% to 100%)"

Intensity of others' emotional experiences (extension)

We extended the target's measure of prevalence (above) with a measure of intensity: "Please try and estimate the emotional intensity for other US American student participants on Prolific taking the survey like you who have experienced this emotion (1 = *Lowest emotional intensity*, and 100 = *Highest emotional intensity*)"

Misestimations of prevalence and intensity

We first calculated the actual prevalence of emotional experiences by counting for each experience the number of participants who indicated having that experience, and then converted that into a percentage. We then calculated for each participant and for each emotional experience the misestimation of the prevalence of that emotional experience as the estimation of prevalence minus actual prevalence of that emotional experience.

We calculated the actual intensity of emotional experience by calculating the average of all self-reported intensity for all those who reported having that experience. We then calculated for each participant and for each emotional experience the misestimation of the intensity as the participant's estimation of intensity of that emotional experience minus the actual intensity of that emotional experience.

This means that for both prevalence and intensity a misestimation score higher than zero indicates an overestimation, and a score lower than zero indicates an underestimation.

Depressive Symptoms

We measured depression using the 10-item short-form version of the Center for Epidemiologic Studies Depression Scale (Radloff, 1977). Respondents indicated the frequency of events in the week prior on items such as "My sleep was restless" and "I had trouble keeping my

mind on what I was doing" (0 = "*Rarely or none of the time (less than 1 day)*"; 3 = "*All of the time (5-7 days)*"; Cole et al., 2004; α = .88).

Brooding/ Rumination

We measured Brooding/ Rumination using the Brooding subscale (5-item) of the Ruminative Responses to Depression Questionnaire (Nolen-Hoeksema, 1991; Treynor, Gonzales & Nolen-Hoeksema, 2003). Participants rated items such as "Why can't I handle things better" and "What am I doing to deserve this" (1 = *almost never*; 4 = *almost always*), scoring from 5 to 20 (α = .87).

Loneliness

We measured participants' loneliness with the 8-item short-form version of the UCLA Loneliness Scale (Hays & DiMatteo, 1987; Russell et al., 1980). Participants rated their agreement with statements such as: "I feel left out" and "I lack companionship" (1 = *Never*; 4 = *Always*; α = .88).

Life Satisfaction

We measured participants' life satisfaction with the 5-item Satisfaction with Life Scale (SWLS; Diener et al., 1985). Participants rated five statements (1 = *Strongly disagree*; 7 = *Strongly agree*), such as "So far I have gotten the important things I want in life" and "The conditions of my life are excellent". (α = .92)

Subjective Happiness

We measured participants' overall happiness using the 4-item Subjective Happiness Scale (SHS). Participants rated four statements on a 7-point Likert scale, with items such as "In

general, I consider myself…" ranging from "not a very happy person" to "a very happy person" (Lyubomirsky & Lepper, 1999) (α = .88).

Number of friends

Participants indicated a rough number of friends with whom they feel comfortable talking to concerning their personal emotional experiences.

Social Comparison Orientation (extension)

We measured social comparison orientation using the 11-item Iowa-Netherlands Comparison Orientation Measure (INCOM) (Gibbons & Buunk, 1999), with items such as "I always like to know what others in a similar situation would do" and "I often like to talk with others about mutual opinions and experiences" (5-point scale; 1 = *Strongly disagree*; 5 = *Strongly agree*; α = .73).

Deviations and replication closeness evaluation

We made several adjustments to the target article's study design. We summarized our deviations with a comparison of study design of the target article with our replication using the LeBel et al. (2018) replication closeness evaluation criteria in Table 5. We categorized the replication as a close to far replication (see "replication closeness evaluation" in the supplementary materials).

Table 5

Classification of the replication based on LeBel et al. (2018)

Design facet	Replication	Details of deviation	Reasons for change
IV construct	Same, with an extension		
DV construct	Similar, with an added extension	We separated prevalence from intensity, by adjusting the self-report measure and adding a separate measure estimating emotional intensity. We added a social comparison orientation individual differences measure.	We proposed that social comparison orientation may interact with the misperception and contribute to differential well-being outcomes.
IV operationalization	Similar		
DV operationalization	Similar	Original: Study 1b conflated emotional prevalence and intensity. Study 3 stripped intensity. Replication: Emotional prevalence and emotional intensity in self were combined into the same question. For others' they were rated as two separate measures - frequency and intensity. Original: "rejected by a boy or girl" includes gender. Replication: adjusted to "rejected by someone" for inclusiveness. We added measures evaluating views on test items that were vague, such as "received a bad grade"	
Population	Similar	The population in our studies were also students, but with a larger more diverse sample from across the United States. Original: $N = 80$ (Study 1b); $N = 104$ (Study 3) Replication: $N = 594$	

Design facet	Replication	Details of deviation	Reasons for change
Procedural details	Different	We added comprehension checks before the rating task to ensure participants understand (1) target (self or others), and (2) dimension (prevalence or intensity)	Ensuring participants read and understood the instructions and to rate accordingly.
		We randomized the self-other rating order and the order of scale measures.	Address order effects
		We added three clarifications questions that inquired about how they perceived subjective test items.	Address possibility of diverging perceptions of our test items
Physical settings	Similar to Study 3	Original: Study 1b: In person; Study 3: Online. Replication: Online.	Target's Studies 1b and 3 had similar designs that can be combined and extended.
Contextual variables	Similar	Original studies: Participants were recruited from a medium-sized West Coast university in the United States. Current replication: Our replication was conducted in 2023 broadly targeting college students across the US.	
Statistical Analyses	Similar with additional analyses that were more suitable and tested robustness	We added participant-level analyses to supplement the item-level analyses in the original article. We also conducted moderation analyses	Misestimations were determined based on estimate-actual comparisons. To examine the effect of social comparison orientation
Replication classification	Close to far replication	Mostly followed the target, yet the self-rating adjustment to measuring both prevalence and intensity warranted a more conservative categorization	

Data analysis strategy

Replication measures

Prevalence estimation error: Replication item-level analyses

We followed the target's analysis and used this as our criteria for a successful replication.

Prevalence estimation error: Item-level one sample t-test for each item

In the target article, the authors first aggregated means of the self-rating of emotional experiences for each of the six negative and six positive items. They then calculated an estimation error comparing participants' estimation of others' emotional experiences against the aggregated mean of all participants' self-ratings. They then conducted a series of 12 one-sample t-tests to examine if participants overestimated by comparing estimation error to 0.

Prevalence estimation error: Item-level one-sample t-test for positive and negative

The authors also conducted two item-level one-sample t-tests, one on the item-level aggregate of the *negative* emotional experiences, and another on the item-level aggregate of the *positive* emotional experiences.

Prevalence estimation error: Item-level independent sample t-test comparing positive and negative

They ran an item-level independent sample t-test of the average estimation error for the negative versus the positive items.

Prevalence estimation error: Extension participant-level analyses

We felt that the replication analysis above could be improved and therefore supplemented the analysis with a participant-level analysis, where we first computed the average for the

negative events for each participant, and then the average for the positive events. We then conducted two participant-level one-sample t-tests, one on the mean of the *negative* emotional experiences, and another on the mean of the *positive* emotional experiences. Finally, we ran a participant-level paired sample t-test of the average estimation error for the negative versus the positive items.

Associations between prevalence estimates of others and well-being: Replication analyses

We followed the target's analyses by computing each participant's average peer-prevalence estimates for negative and positive emotional experiences. Thereafter, we conducted a linear regression analysis to examine the associations between emotional estimation error and the well-being indicators such as depressive symptoms, rumination level, life satisfaction, loneliness, happiness level, and the number of confidants.

Extension measures

Intensity estimation error

We added intensity measures and conducted all the analyses for prevalence above also for intensity, for both item-level and participant-level. We also conducted similar analyses for associations with well-being measures.

Social comparison orientation

We added the social comparison orientation (SCO) measure to all the correlational and regression analyses detailed above for both prevalence and intensity, with all the other well-being measures. We also conducted regression interaction analyses to examine whether SCO interacts with estimation error in predicting well-being measures.

Outliers and exclusions

We followed the pre-registered plan to only include responses from participants who completed the entire questionnaire with no further exclusions. We also pre-registered that in case we failed to find support for the estimation hypotheses in our replication of the target article, we would supplement our analyses by rerunning the analyses with exclusion and stricter alpha to account for multiple analyses (alpha = .005). We found support for misestimation effects (see below) using the full sample ($N = 594$) and therefore did not proceed to conduct or report the analyses with exclusions.

Results

Prevalence Estimate Errors (replication)

We summarized descriptives in Tables 6 (prevalence) and 7 (intensity), and correlations in Table 8.

Consistent with Hypothesis 2a and the original findings, we found support for an underestimation of others' negative emotional experiences. Mirroring the analyses conducted in the original article, we computed the estimation error for each item of positive and negative experiences by comparing the self-rating with ratings for their peers. We conducted an item-level one-sample t-test on the estimation errors to examine if they differed from 0. We found support for the expected underestimation of negative ($t(5) = -4.30$, $p = .008$, $d = -1.76$, 95% CI [-3.06, -0.41]), yet also support for an unexpected underestimation of positive experiences ($t(5) = -3.67$, $p = .014$, $d = -1.50$, 95% CI [-2.67, -0.26]).

We further extended our analyses to a series of participant-level one-sample t-tests on the estimation error of emotional experiences, summarized in Table 6. Overall, we found support for

participants' underestimation of all six negative emotional experiences and of five of the six positive emotional experiences. The participant-level analyses mirrored that of the item-level analyses, with support for the expected underestimation of negative emotional experiences ($M =$ 23.56; $t(593) = -35.02$, $p < .001$, $d = -1.44$, 95% CI [-1.55, -1.32]), yet again with the unexpected underestimation of positive emotional experiences ($M = 15.97$; $t(593) = -24.17$, $p < .001$, $d = -0.99$, 95% CI [-1.09, -0.89]).

Partially consistent with Hypothesis 2b, the underestimation of negative experiences was stronger than positive experiences, yet only for the participant-level analysis. Specifically, for the item-level analyses, we ran an independent samples t-test comparing negative and positive emotional experiences and did not find support for Hypothesis 2b in estimation error ($t(10) =$ 0.31, $p = .31$, $d = -0.62$, 95% CI [-1.79, 0.60]). We anticipated this in advance, given that the item-level analysis had too few items and therefore power to detect such differences, which is why we planned and pre-registered to also run participant-level analyses, which we felt were more appropriate and accurate. The participant-level paired-samples t-test analysis allowed us to find support a larger estimation error for negative compared to positive experiences ($t(593) = -11.90$, $p < .001$, $d = -0.49$, 95% CI [-0.57, -0.40]).

Intensity Estimate Errors (extension)

To test whether underestimation of emotional experiences also extends to emotional intensity, we conducted similar analyses on the emotional intensity measures. We began with item-level one-sample t-tests and found support Hypothesis 2d with an overestimation of positive experiences ($t(5) = 2.97$, $p = .031$, $d = 1.21$, 95% CI [.099, 2.26]), yet contrary to our expectations in Hypothesis 2c, we also found an overestimation for negative experiences ($t(5) = 3.51$, $p = .017$, $d = 1.43$, 95% CI [0.23, 2.58]).

The participant-level analyses showed a similar trend, as we found support for the expected overestimation of positive emotional experiences ($M = 5.60$; $t(593) = 8.45$, $p < .001$, $d = 0.35$, 95% CI [0.26, 0.43]), and again the unexpected overestimation of negative emotional experiences ($M = 12.54$; $t(593) = 19.42$, $p < .001$, $d = 0.80$, 95% CI [0.70, 0.89]). We ran one-sample t-tests for each of the experiences and found that participants overestimated five out of the six negative emotional experiences and five out of the six positive emotional experiences.

Mirroring the target's analyses for item-level prevalence comparing positive and negative, we ran the same analysis for intensity estimates and found no signal for difference in overestimation error between negative and positive emotional experiences, though large effect size ($t(10) = 1.72$, $p = .117$, $d = 0.99$, 95% CI [-0.24, 2.18]). Again, this is most likely due to the small number of items which a power analysis shows requires an unreasonably large effect in order to be detectable with null hypothesis significance testing. We supplemented the item-level analysis by conducting a participant level paired sample t-test, which was far better powered, and indeed found support for stronger overestimation for negative experiences than for positive experiences ($t(593) = 10.63$, $p < .001$, $d = 0.44$ 95% CI [0.35, 0.52]).

Prevalence estimates associations with well-being (replication)

We summarized the Pearson's correlations in Table 8 and the regression model findings predicting well-being from negative and positive prevalence estimates in Table 9 (comparing with the target's).

Inconsistent with and opposite to Hypothesis 4, prevalence estimations of negative emotional experiences were negatively associated with well-being, as indicated by a negative association with life satisfaction ($\beta = -.23$, $t(591) = -4.87$, $p < .001$), and subjective happiness ($\beta =$

-.26, $t(591) = -5.59$, $p < .001$), and a positive association with loneliness ($\beta = .33$, $t(591) = 7.04$, $p < .001$), brooding ($\beta = .37$, $t(591) = 8.08$, $p < .001$), depressive symptoms ($\beta = .41$, $t(591) = 9.01$, $p < .001$). On the other hand, positive prevalence estimates were negatively associated with loneliness ($\beta = -.25$, $t(591) = -5.40$, $p < .001$), and depressive symptoms ($\beta = -.22$, $t(591) = -4.69$, $p < .001$), and positively associated with life satisfaction ($\beta = .25$, $t(591) = 5.35$, $p < .001$), subjective happiness ($\beta = .32$, $t(591) = 6.73$, $p < .001$), and the number of confidants ($\beta = .19$, $t(591) = 3.98$, $p < .001$). We conclude that our findings are in the opposite direction to that reported in the target article, for both negative and positive emotional experiences prevalence estimates.

Table 6

Prevalence of emotional experiences (replication and extension): One-sample t-tests of estimation error

Experiences	Actual prevalence average	Prevalence estimate mean	Error mean	Error SD	t-stat	P	Cohen's d and CI	Interpretation	t-stat	df	p	Cohen's d and CI
Negative experiences												
Had fight/argument	71.04	47.41	-23.63	23.54	-24.47	< .001	-1.00 [-1.10, -0.90]	Signal; same direction	/	/	/	/
Thought about distant friends/ family	78.79	45.76	-33.02	22.61	-35.60	< .001	-1.46 [-1.58, -1.34]	Signal; same direction	/	/	/	/
Thought about enormous workload	86.36	62.14	-24.22	22.29	-26.49	< .001	-1.09 [-1.19, -0.99]	Signal; same direction	/	/	/	/
Was rejected by boy/girl	46.13	36.04	-10.09	22.11	-11.12	< .001	-0.46 [-0.54, -0.37]	Signal; same direction	/	/	/	/
Received low grade	50.17	42.60	-7.57	23.60	-7.82	< .001	-0.32 [-0.40, 0.24]	Signal; same direction	/	/	/	/
Thought about bad personal health habits	85.35	42.51	-42.85	23.02	-45.36	< .001	-1.86 [-1.99, -1.73]	Signal; same direction	/	/	/	/
Overall negative	**69.64**	**46.08**	**-23.56**	**16.40**	**-35.02**	**< .001**	**-1.44 [-1.55, -1.32]**	**Signal; same direction**	**-4.30**	**5**	**.008**	**-1.76 [-3.06, -0.41]**
Overall negative: Target article (Study 3)			**-21.4**	**8.7**	**-**	**-**		**-**	**-5.99**	**5**	**< .01**	**-2.68 [-4.65, -0.68]**
Positive experiences												
Received high grade	75.08	54.84	-20.35	24.35	-20.27	< .001	-0.83 [0.92, 0.74]	Unexpected signal; same direction	/	/	/	/
Attended fun party	51.35	52.36	1.01	24.04	1.03	0.305	0.04 [-0.04, 0.12]	Unexpected no signal;	/	/	/	/
Participated in athletics	54.04	39.45	-14.59	22.64	-15.70	< .001	-0.64 [-0.73, 0.56]	Signal; opposite direction	/	/	/	/
Went out with friends	77.27	65.61	-11.66	19.63	-14.48	< .001	-0.59 [-0.68, -0.51]	Signal; opposite direction	/	/	/	/
Talked to distant friends/ family	72.90	53.68	-19.22	22.47	-20.84	< .001	-0.86 [-0.95, 0.76]	Signal; opposite direction	/	/	/	/
Had great meal	94.44	63.31	-31.22	22.00	-34.49	< .001	-1.42 [-1.53, -1.30]	Unexpected signal; same direction	/	/	/	/
Overall positive	**70.85**	**54.87**	**-15.97**	**16.10**	**-24.17**	**< .001**	**-0.99 [-1.09, -0.89]**	**Unexpected signal**	**-3.67**	**5**	**.014**	**-1.50 [-2.67, -0.26]**
Overall positive: Target article (Study 3)			**3.8**	**8.7**	**-**	**-**	**-**	**-**	**1.06**	**5**	**n.s**	**0.48 [-0.48, 1.38]**
Overall positive and negative combined	70.24	50.48	-19.77	14.27	-33.75	< .001	-1.38 [-1.50, -1.27]		-5.61	11	< .001	-1.62 [-2.48, -.73]

Note. One-sample t-tests, *N* = 594, participant level *df* = 593. CI 95% confidence intervals. The interpretation of the outcome is based on LeBel et al. (2019), we summarized whether we found a signal (p < alpha), whether the signal or lack of was expected, and whether it was in the same direction as that of the target's per the singular item. We did not summarize effect size consistency (target's effect within replication confidence intervals) given that we conducted analyses on participant level whereas the target conducted analysis on item-level.

Table 7

Intensity of emotional experiences [Extension]: One-sample t-tests of estimation error

Experiences	Participant-level analyses							Item-level analyses			
	Actual intensity average	Intensity estimate mean	Error mean	Error SD	*t*-stat	*p*	Cohen's *d* and CI	t-stat	df	*p*	Cohen's d and CI
Negative experiences											
Had fight/argument	42.63	59.64	17.01	22.27	18.62	< .001	0.76 [0.67, 0.86]	/	/	/	/
Thought about distant friends/ family	42.65	48.48	5.84	21.48	6.62	< .001	0.27 [0.19, 0.35]	/	/	/	/
Thought about enormous workload	50.81	60.44	9.63	22.13	10.60	< .001	0.43 [0.35, 0.52]	/	/	/	/
Was rejected by boy/girl	33.00	58.16	25.17	23.94	25.62	< .001	1.05 [0.95, 1.15]	/	/	/	/
Received low grade	37.05	53.65	16.60	22.46	18.01	< .001	0.74 [0.65, 0.83]	/	/	/	/
Thought about bad personal health habits	44.07	45.07	1.01	21.93	1.12	= .264	0.05 [-0.03, 0.13]	/	/	/	/
Overall negative	**41.70**	**54.24**	**12.54**	**15.73**	**19.42**	**< .001**	**0.80 [0.70, 0.89]**	**3.51**	**5**	**.017**	**1.43 [.23, 2.58]**
Positive experiences											
Received high grade	56.07	60.56	4.50	22.39	4.90	< .001	0.20 [0.12, 0.28]	/	/	/	/
Attended fun party	50.63	62.65	12.03	21.35	13.73	< .001	0.56 [0.48, 0.65]	/	/	/	/
Participated in athletics	44.95	49.35	4.40	22.16	4.84	< .001	0.20 [0.12, 0.28]	/	/	/	/
Went out with friends	59.11	67.65	8.54	19.47	10.69	< .001	0.44 [0.35, 0.52]	/	/	/	/
Talked to distant friends/ family	52.03	57.96	5.93	20.29	7.12	< .001	0.29 [0.21, 0.37]	/	/	/	/
Had great meal	58.80	57.03	-1.77	23.40	-1.85	= .065	-0.08 [-0.16, 0.004]	/	/	/	/
Overall positive	**53.60**	**59.20**	**5.60**	**16.16**	**8.45**	**< .001**	**0.35 [0.26, 0.43]**	**2.97**	**5**	**.031**	**1.21 [.099, 2.26]**
Overall: positive and negative combined	47.65	56.72	9.07	13.83	15.99	< .001	0.66 [0.57, 0.74]	4.14	11	.002	1.20 [0.43, 1.90]

Note. One-sample t-tests, *N* = 594, participant level *df* = 593. CI 95% confidence intervals.

Table 8

Correlations between prevalence and intensity estimates with well-being and social comparison measures

Variable	M	SD	alpha	1	2	3	4	5	6	7	8	9	10	11
1- Prevalence estimates for *negative* emotions	46.1	16.4	-											
2 - Prevalence estimates for *positive* emotions	54.9	16.1	-	.54*** [.48, .60]										
3 - Prevalence estimates *overall*	50.47	14.27	-	.88*** [.86, .90]	.88*** [.86, .89]									
4 - Intensity estimates of *negative* emotions (extension)	54.2	15.7	-	.58*** [.53, .63]	.35*** [.28, .42]	.53*** [.47, .59]								
5 - Intensity estimates of *positive* emotions (extension)	59.2	16.2	-	.32*** [.25, .39]	.63*** [.58, .67]	.54*** [.48, .59]	.50*** [.44, .56]							
6 - Intensity estimation error *overall* (extension)	9.07	13.83	-	.52*** [.46, .58]	.57*** [.51, .62]	.62*** [.47, .67]	.86*** [.84, .88]	.87*** [.85, .89]						
7 - Social Comparison Orientation (extension)	35.30	6.27	.67	.21*** [.13, .28]	.04 [-.04, .12]	.14*** [.06, .22]	.14*** [.06, .22]	.08 [-.01, .16]	.12** [.04, .20]					
8 - Loneliness	18.36	5.97	.88	.19*** [-.11, .27]	-.07 [-.15, .01]	.07 [-.01, .15]	.18*** [.10, .26]	-.07 [-.15, .01]	.06 [-.02, .14]	.28*** [.20, .35]				
9 - Brooding	11.44	3.95	.87	.30*** [.22, .37]	.06 [-.02, .14]	.20*** [.13, .28]	.17*** [.09, .25]	.02 [-.07, .10]	.11** [.03, .19]	.45*** [.38, .51]	.56*** [.51, .62]			
10 - Depression	10.28	6.60	.88	.30*** [.22, .37]	.01 [-.07, .09]	.18*** [.10, .25]	.19*** [.11, .27]	-.05 [-.13, .03]	.08* [.00, .16]	.32*** [.25, .39]	.72*** [.68, .75]	.69*** [.64, .73]		
11 - Life satisfaction	20.84	7.75	.92	-.09* [-.17, -.13]	.13* [.05, .21]	.02 [-.06, .10]	-.05 [-.13, .03]	.15*** [.07, .23]	.06 [-.02, .14]	-.17*** [-.25, -.09]	-.54*** [-.59, -.48]	-.45*** [-.52, -.39]	-.55*** [-.60, -.49]	
12 - Subjective Happiness	17.15	5.55	.88	-.09* [-.17, -.11]	.17*** [.09, .25]	.05 [-.04, .13]	-.10* [-.18, -.02]	.20*** [.12, .28]	.06 [-.02, .14]	-.18*** [-.26, -.11]	-.65*** [-.69, -.60]	-.47*** [-.53, -.40]	-.65*** [-.69, -.60]	.65*** [.61, .70]

Note. * $p < .05$; ** $p < .01$; *** $p < .001$. df = 592.

Table 9

Prevalence estimates regression coefficients with outcomes: comparison between Jordan et al.

(2011)'s Study 3 versus replication

Measure	Jordan et al. (2011)				Replication			Interpretation
	Cronbach's α	β for negative prevalence estimate	β for positive prevalence estimate	R^2	β for negative prevalence estimate	β for positive prevalence estimate	R^2	
Negative								
Loneliness	.81	-.30**	.08	.08	.33***	-.25***	.08	Signal; Inconsistent opposite
Rumination/brooding	.72	-.28**	.17	.08	.37***	-.15**	.10	Signal; Inconsistent opposite
Depressive symptoms	.73	.00	-.02	.00	.41***	-.22***	.12	Signal; Inconsistent opposite
Positive								
Satisfaction with life	.84	.23*	-.37***	.15	-.23***	.25***	.05	Signal; Inconsistent opposite
Subjective Happiness	.80	.19	-.16	.05	-.26***	.32***	.08	Signal; Inconsistent opposite
Number of confidants	/	N/A	N/A	N/A	-.11*	.19***	.03	
Social orientation scale (extension)	/	/	/	/	.26***	-.10*	.05	Supported

Note. Linear regression, $N = 594$. The "original" column was adopted from Jordan et al. (2011), pp. 130.
*$p < .05$. **$p < .01$. ***$p < .001$. N/A = not reported. The interpretation of the outcome is based on LeBel et al. (2019).

Intensity estimates associations with well-being (extension)

We conducted the same analyses on intensity estimates as the ones reported above for prevalence, and summarized those in Tables 8 and 10. We found a similar pattern, with support for a positive association between negative emotional experiences intensity estimates and loneliness ($\beta = .29$, $t(591) = 6.21$, $p < .001$), brooding ($\beta = .22$, $t(591) = 4.74$, $p < .001$), and depressive symptoms ($\beta = .29$, $t(591) = 6.31$, $p < .001$), and negative association with life satisfaction ($\beta = -.18$, $t(591) = -3.79$, $p < .001$), and subjective happiness ($\beta = -.28$, $t(591) = -6.07$, $p < .001$).

On the other hand, we found that positive emotional experiences intensity had a negative association with loneliness ($\beta = -.21$, $t(591) = -4.63$, $p < .001$) and depressive symptoms ($\beta = -.20$, $t(591) = -4.23$, $p < .001$).

Complementary analysis: Self-reports associations with well-being (exploratory extension)

We explored the associations between self-reports of emotional experiences, and well-being measures. We found that reporting more of the negative emotional experiences was positively correlated with loneliness ($r = .16$ [.08, .23]), brooding ($r = .25$ [.17, .32]), and depression ($r = .26$ [.19, .34]), and that reporting more the listed positive emotional experiences was positively correlated with life satisfaction ($r = .22$ [.14, .29]) and happiness ($r = .27$ [.20, .35]) and negatively associated with loneliness ($r = -.16$ [-.23, -.07]; all $p < .001$).

Complementary analysis: interaction between self and others in predicting well-being (exploratory extension)

We also explored interactions between one's own negative experiences and estimates of others' negative experiences in predicting well-being, and found support for an interaction for both loneliness, brooding, and depression, such that the positive association between prevalence estimates and negative factors of well-being was stronger the less negative experiences one had. We also found support for an interaction with happiness and wellbeing, such that the negative association between prevalence estimates and positive factors of well-being was stronger the less negative experiences one had. We did not find such interactions for the positive experiences.

Table 10

Intensity estimates regression coefficients with outcomes (extension)

Measure	β for negative prevalence estimate	β for positive prevalence estimate	R^2
Negative well-being			
Loneliness	.29***	-.21***	.07
Rumination/brooding	.22***	-.10*	.04
Depressive symptoms	.29***	-.20***	.07
Positive well-being			
Satisfaction with life	-.18***	.24***	.05
Subjective Happiness	-.28***	.34***	.10
Number of confidants	-.07	.15**	.02
Social orientation scale (extension)	.13**	.01	.02

Note. Linear regression, $N = 594$. *$p < .05$; **$p < .01$; ***$p < .001$.

External analysis: Suppression using target's analyses

When consulting with external expert reviewers to examine the possible explanation for our associations being in the opposite direction from that of the target article, a reviewer was kind to review the target article's and our results. In his analyses (shared on https://osf.io/zy5qa/), he pointed out that there might be suppression in the regression analyses (Sharpe & Roberts, 1997; Thompson & Levine, 1997) that were conducted by the target article and that we repeated in our replication. His conclusion was that the regression analyses should at best be interpreted with caution, and that interpretations should be focused on the raw correlation effects, which were weaker yet in the same direction (see Table 8).

We reached out to the original authors of the target article, and received the reply that the raw data, analysis code, and correlations for the target article's studies are not available. However, the reviewer, we, and the original authors agreed that suppression alone is not likely to explain the complete reversal of the pattern of effects.

Social Comparison Orientation (exploratory extension)

We conducted Pearson's correlation tests to examine the associations between social comparison orientation and well-being variables, and summarized them in Table 8. We found support for Hypotheses 6a to 6c, in that social comparison orientation was positively associated with depressive symptoms ($r(592) = .32$, 95% CI [.25, .39], $p < .001$), loneliness ($r(592) = .28$, 95% CI [.20, .35], $p < .001$), and rumination ($r(592) = .45$, 95% CI [.38, .51], $p < .001$). We also found support for Hypotheses 6d and 6e that social comparison orientation was negatively associated with subjective happiness ($r(592) = -.18$, 95% CI [-.26, -.11], $p < .001$), and life satisfaction ($r(592) = -.17$, 95% CI [-.25, -.09], $p < .001$).

We also examined interactions with the misestimation of emotional experiences in predicting psychological well-being. We failed to find support for social comparison orientation as moderating the associations between misestimations and well-being measures. We summarized the results with plots of the moderation analysis in the supplementary materials. Given the many analyses conducted and many finding no support, we therefore only summarized interactions that were supported and documented below $p < .05$, yet we strongly caution against over-interpreting those and recommend focusing on the much stronger and clearer main effects.

Discussion

We conducted a pre-registered replication of Studies 1b and 3 from Jordan et al. (2011) and tested their theoretical framework, measurement, and analysis strategy with a larger sample. We also went beyond the target article by adding additional tests that seem like a bit fit for the experimental design, and by adding extensions testing for the generalizability to emotional experience intensity and examining social comparison orientation.

Our results were mixed, and we summarized a comparison of our findings to that of the target in Tables 6 and 9. We concluded a successful replication only regarding people's underestimation of the prevalence of others' negative emotions, yet with an unexpected underestimation of the prevalence of positive emotions. In addition, we only observed differences between the underestimation of positive and negative emotional experiences when conducting the better-powered participant-level analysis, yet not when repeating that target's item-level analysis, likely due to the small number of items and the analysis being underpowered.

Most surprising was that we found opposite effects to the target's theory and findings regarding the associations between prevalence estimates and psychological well-being factors. Based on the target's findings we expected higher estimations for prevalence of negative emotions to be positively associated with higher well-being, and instead we found support for a negative relationship. We discuss possible explanations for the mixed findings, followed by a discussion of limitations and suggestions for future research directions.

Replication: Prevalence

We conclude mixed findings concerning our replication of the systematic misperception of others' emotional experiences. We found that: (1) people underestimated the prevalence of others' negative and positive emotional experiences, (2) underestimation errors were stronger for negative experiences than for positive experiences, (3) prevalence estimation of others' negative emotions was positively associated with loneliness, rumination, depressive symptoms, and negatively associated with life satisfaction and subjective happiness, (4) prevalence estimation of others' positive emotions was positively associated with life satisfaction and subjective happiness and negatively associated with loneliness, rumination, and depressive symptoms.

Jordan et al. (2011) argued that underestimation was mainly about negative emotional events, yet instead, we found that participants also underestimated positive emotional experiences, albeit to a lesser extent than they did for negative events. This supports a needed reframing of their null hypothesis that we suggested in our pre-registration (Hypothesis 2b; see Table 2), that instead of null effects for positive emotions, the hypothesis could be that the underestimation of positive events is weaker than that of negative events. We note, however, that there would still be an unexplained inconsistency with the target article's Study 2, which

extended the idea from Study 1, and using a preexisting dataset from Srivastava et al. (2009) showed very large differences in misestimation between negative and positive emotions, such that negative emotions were underestimated (8 out of 9 emotions), compared to the an overestimation of positive emotions (7 out of 8 emotions). We therefore see much value in conducting also a follow-up replication of the target's Study 2 with possible theoretical and empirical extensions that would try and resolve the differences in results.

Prevalence estimations and well-being

We failed to find support for the findings regarding correlates of prevalence estimates with well-being measures of loneliness, rumination, depressive symptoms, life satisfaction, and subjective happiness. Instead, we found support for the opposite effects of those reported in the target article.

It is difficult to resolve the inconsistent findings regarding the associations between prevalence estimations and well-being. We reached out to the target article's authors to consult with them regarding the diverging findings (May, 2023), and we were unable to identify the reason for the divergence.

 We note that in our view there was a misalignment between the theoretical framework and the analyses performed in the target article that may at least partly account for the contradictory replication findings. Jordan et al. (2011) argued that *feeling alone* in negative emotional experiences may lead to the feeling that negative emotions are less common and subsequently pathologize their experiences, causing negative psychological impact. Consistent with the target's findings, Whillans and colleagues (2017) found that first year arrivals who perceived themselves as less socially connected than their peers reported lower belonging and

well-being. However, the methods in the target article are not fully aligned with the testing of their main theory, since the main argument seems to refer to a comparison between self and others' emotional experiences, which differs from the main effect correlational analyses that they reported to test their hypotheses. The original analyses may have stemmed from an assumption that the rating person had experienced the negative emotional events before and so perceiving others as also experiencing those negative events helped them feel less alone. However, an alternative scenario might be that the rating person had not experienced the negative emotional event before, and so thinking that negative experiences were common contributed to them feeling more alone. In our replication analyses we first simply followed their methodology and data analysis strategy, yet a more suitable data analysis would have been to examine the interaction between one's own experiences and one's perception of others and its association with factors like loneliness. We added exploratory analyses that indeed suggest that self-reports and estimates of others' negative emotional experiences may interact in predicting well-being factors, such that having fewer negative experiences makes viewing more others as having negative experiences predict one's own lower well-being.

We therefore offer several suggested hypotheses to be confirmed in future studies: 1) feeling alone in emotional experiences is negatively associated with well-being, 2) for those who have experienced negative emotions, a lower prevalence estimate makes them feel as struggling alone and hence be associated with poorer well-being, and 3) for those who have not experienced negative emotions, a higher prevalence estimate of negative emotions is associated with poorer well-being. Given the correlational methods, the causal direction remains unclear and future research may try to further explore causality in the interplay between prevalence estimations and well-being.

Extension: Intensity

We ran extensions examining if the underestimations extend to an underestimation of others' emotional *intensity*. Our findings showed that: (1) people overestimate the intensity of others' negative emotional experiences, (2) people overestimate the intensity of others' positive emotional experiences, but to a lesser extent than for negative emotions. This diverged from our initial predictions that misestimations of intensity would differ for negative versus positive, and should therefore be subjected to further confirmation with additional research.

Challenging and reframing misestimation: Prevalence and intensity, positive and negative

Our findings challenge the target article's findings in several important ways. First, the effects seem to encompass both positive and negative emotional events, yet with stronger prevalence underestimations for negative than for positive emotional events. We also found that misestimations of others' emotions extend to misestimations of others' emotional intensity, yet, crucially, in opposite directions. Our findings suggest that people underestimate prevalence yet at the same time overestimate intensity. If that is indeed the case, then the story shifts from people not being sensitive enough to others' positive emotions to people sensing fewer but stronger intensity instances of both negative and positive emotions.

One direction is focusing on the broadcasting side of interpersonal exchanges, on how people exhibit their emotional experiences. It could be that estimations are accurate and effectively capture what people indeed exhibit - that people suppress or are able to hide low intensity emotional events (Srivastava et al., 2009) yet are less able to suppress their high intensity emotional events.

A different possibility all together is focusing on the receiving side, in that it might not be at whether emotional experiences occur in solitude or how they are exhibited, but rather more about how we sense, code, remember, and recall social information about emotional experiences, in that people tend to focus on, respond to, and remember the strongest emotional events. People rely on vivid memories of past expressions of emotional experiences (Doré et al., 2016), which tend to be events that are emotionally intense (Yonelinas & Ritchey, 2015). Hence, the most salient expressions may color their evaluations of others' experiences. While they may not recall many instances, leading to an underestimate of prevalence, the ones that are recalled are the most vivid ones, leading to an overestimation of intensity.

Extension: Social comparison orientation

We found support for associations between social comparison orientation, misestimations, and well-being: (1) social comparison orientation was positively associated with depressive symptoms level, loneliness, and brooding, (2) social comparison orientation was negatively associated with life satisfaction and subjective happiness, and (3) social comparison orientation was positively associated with prevalence estimates and intensity estimates overall.

These associations are in line with the wealth of literature suggesting that social comparison orientation, which is mostly upward social comparison (Festinger, 1954), created a discrepancy between the ideal self-presented to others and the real self (Yu & Kim, 2020). This induced a sense of inferiority and distress as these comparisons maintain or even exacerbate negative self-evaluations (Vogel et al., 2015). As a result, these negative evaluations about oneself were found to cause detrimental effects on well-being characterized by lower life satisfaction and lower subjective well-being (Verduyn et al., 2015).

We predicted yet found no support for the idea that social comparison orientation interacts with the association between misestimations and psychological well-being. This may be suggestive of the story being less about comparisons between self and others and more about how people communicate or receive social-emotional information, though null effects should be interpreted with caution and humility. Future research can further contrast social comparisons with social attention/awareness and/or empathy as impacting emotional event misestimations.

Limitations and future directions

We note several limitations in our current replication, which may partly explain some of our diverging findings. We noted several weaknesses we spotted in the target's analyses and adjustments we made to address those (see Table 5 for the summary of deviations), and so it is possible that one of those affected the findings. For example, we followed the target's method and their items a as closely and yet some items were specific about an emotion regarding a specific event (e.g. feeling sad because of a low grade), and it is possible that the estimate of the frequency of the emotions was based on estimates of the frequency of the event and that for the described events there were other emotions that were more relevant than those that the authors had in mind (e.g., feeling shame rather than sadness over a low grade). To avoid that, future research may aim to separate context from emotions, to help gain a more accurate understanding of the cause for misestimations.

We note that we adopted the same emotional experiences items that appeared in the original article, trying to stay as close as possible to the setting of the original article. Yet, the procedure in the target article included a study aiming to generate the items, meant to adjust the items to be as relevant as possible to the target sample. Though we were careful to run the study

with a sample of students that would find the experiences listed relevant, it could be that not following the whole procedure may have affected the findings somehow. We do not think this to be a major issue, given the strong support for the targets' findings regarding prevalence underestimation, which suggests these were suitable.

Our sample was similar to the target article's sample with US American students, yet with a more diverse and more heterogeneous population recruited online on Prolific compared to the target article's sample which was recruited from a single school. It is possible that this change may have impacted our findings in some way and that there are some differences (socio-cultural, economics, etc.), which would on the one hand show that the consistent findings are generalizable yet that the heterogeneity may have led to some of the inconsistent findings. We believe that research should aim for more diverse heterogenous samples to go beyond a singular context, sample, or point in time, in order to maximize impact and practical use. Future research is needed to examine the generalizability of the phenomenon and retest our and the target article's methods in other contexts.

Conclusion

Overall, we found only partial support for the findings of Jordan et al. (2011) research regarding misestimations of the prevalence of others' emotional experiences. Consistent with the original article, we found support for an underestimation of the prevalence of others' negative and positive emotional experiences. However, inconsistent with the target article, we found that the estimation of negative emotions prevalence was positively associated with loneliness, rumination, and depressive symptoms, and negatively associated with life satisfaction and subjective happiness. On the other hand, the estimation of positive emotions prevalence was positively associated with life satisfaction and subjective happiness and negatively associated with loneliness, rumination, and depressive symptoms. We also ran an extension examining estimation of others' emotional intensity, and unexpectedly found that people tended to overestimate others' positive and negative emotional intensity. In another extension, we found social comparison orientation as associated with misestimation and well-being yet does not moderate the link between prevalence estimations and well-being.

References

Alfasi, Y. (2019). The grass is always greener on my Friends' profiles: The effect of Facebook social comparison on state self-esteem and depression. *Personality and Individual Differences*, *147*, 111-117. https://doi.org/10.1016/j.paid.2019.04.032

Allport, F. H. (1924). *Social psychology*. Cambridge, MA: Riverside Press.

Bjerring, J. C., Hansen, J. U., & Pedersen, N. J. L. L. (2014). On the rationality of pluralistic ignorance. *Synthese*, *191*, 2445-2470. https://doi.org/10.1007/s11229-014-0434-1

Burke, M., & Kraut, R. (2013, February). Using Facebook after losing a job: Differential benefits of strong and weak ties. In *Proceedings of the 2013 conference on Computer supported cooperative work* (pp. 1419-1430). https://doi.org/10.1145/2441776.2441936

Cole, J. C., Rabin, A. S., Smith, T. L., & Kaufman, A. S. (2004). Development and validation of a Rasch-derived CES-D short form. *Psychological assessment*, *16*(4), 360. https://dx.doi.org/10.1037/1040-3590.16.4.360

Darley, J. M., & Latané, B. (1968). Bystander intervention in emergencies: diffusion of responsibility. *Journal of Personality and Social Psychology*, *8*(4), 377-383.

Diener, E. D., Emmons, R. A., Larsen, R. J., & Griffin, S. (1985). The satisfaction with life scale. *Journal of personality assessment*, *49*(1), 71-75. https://doi.org/10.1207/s15327752jpa4901_13

Doré, B. P., Meksin, R., Mather, M., Hirst, W., & Ochsner, K. N. (2016). Highly accurate prediction of emotions surrounding the attacks of September 11, 2001 over 1-, 2-, and 7-year prediction intervals. *Journal of Experimental Psychology: General*, *145*(6), 788. https://doi.org/10.1037/xge0000168

Ekman, P., & Friesen, W. V. (1969). The repertoire of nonverbal behavior: Categories, origins, usage, and coding. *semiotica*, *1*(1), 49-98. https://doi.org/10.1515/semi.1969.1.1.49

Feldman, G. (2023). Registered Report Stage 1 manuscript template.

https://doi.org/10.17605/OSF.IO/YQXTP

Festinger, L. (1954). A theory of social comparison processes. *Human Relations, 7,* 117–140.

https://doi.org/10.1177/001872675400700202

Geiger, N., & Swim, J. K. (2016). Climate of silence: Pluralistic ignorance as a barrier to climate

change discussion. *Journal of Environmental Psychology*, *47*, 79-90.

https://doi.org/10.1016/j.jenvp.2016.05.002

Gibbons, F. X., & Buunk, B. P. (1999). Individual differences in social comparison: development

of a scale of social comparison orientation. *Journal of personality and social psychology*,

76(1), 129. https://doi.org/10.1037/0022-3514.76.1.129

Hays, R. D., & DiMatteo, M. R. (1987). A short-form measure of loneliness. *Journal of*

personality assessment, *51*(1), 69-81. https://doi.org/10.1207/s15327752jpa5101_6

Jané, M., Xiao, Q., Yeung, S., Ben-Shachar, M. S., Caldwell, A., Cousineau, D., Dunleavy, D. J.,

Elsherif, M., Johnson, B., Moreau, D., Riesthuis, P., Röseler, L., Steele, J., Vieira, F.,

Zloteanu, M., & Feldman, G. (2024). Guide to Effect Sizes and Confidence Intervals.

http://dx.doi.org/10.17605/OSF.IO/D8C4G

Jordan, A. H., Monin, B., Dweck, C. S., Lovett, B. J., John, O. P., & Gross, J. J. (2011). Misery

has more company than people think: Underestimating the prevalence of others' negative

emotions. *Personality and Social Psychology Bulletin*, *37*(1), 120-135.

Korte, C. (1972). Pluralistic ignorance about student radicalism. *Sociometry*, 576-587.

https://doi.org/10.2307/2786534

Larson, R., Csikszentmihalyi, M., & Graef, R. (1982). Time alone in daily experience: Loneliness

or renewal. *Loneliness: A sourcebook of current theory, research and therapy*, 40-53.

LeBel, E. P., McCarthy, R. J., Earp, B. D., Elson, M., & Vanpaemel, W. (2018). A unified framework to quantify the credibility of scientific findings. *Advances in Methods and Practices in Psychological Science*, *1*, 389-402.

LeBel, E. P., Vanpaemel, W., Cheung, I., & Campbell, L. (2019). A brief guide to evaluate replications. *Meta-Psychology*, 3, 1-9. DOI: 10.15626/MP.2018.843

Lee J. K. (2022). The effects of social comparison orientation on psychological well-being in social networking sites: Serial mediation of perceived social support and self-esteem. *Current psychology (New Brunswick, N.J.)*, *41*(9), 6247–6259. https://doi.org/10.1007/s12144-020-01114-3

Leviston, Z., Walker, I., & Morwinski, S. (2013). Your opinion on climate change might not be as common as you think. *Nature Climate Change*, *3*(4), 334-337. https://doi.org/10.1038/nclimate1743

Li, M. (2016). Pluralistic ignorance in Sino–Hong Kong conflicts: the perception of mainland Chinese people living in Hong Kong. *Chinese Journal of Communication*, *9*(2), 153-172. https://doi.org/10.1080/17544750.2015.1091021

Liu, P., He, J., & Li, A. (2019). Upward social comparison on social network sites and impulse buying: A moderated mediation model of negative affect and rumination. *Computers in Human Behavior*, *96*, 133-140. https://doi.org/10.1016/j.chb.2019.02.003

Lyubomirsky, S., & Lepper, H. S. (1999). A measure of subjective happiness: Preliminary reliability and construct validation. *Social indicators research*, *46*, 137-155. https://doi.org/10.1023/A:1006824100041

Miller, D. T., & Prentice, D. A. (1994). Collective errors and errors about the collective. *Personality and Social Psychology Bulletin*, *20*(5), 541-550.

Miyajima, T., & Yamaguchi, H. (2017). I want to but I won't: Pluralistic ignorance inhibits intentions to take paternity leave in Japan. *Frontiers in Psychology*, *8*, 1508. https://doi.org/10.3389/fpsyg.2017.01508

Nolen-Hoeksema, S. (1991). Responses to depression and their effects on the duration of depressive episodes. *Journal of abnormal psychology*, *100*(4), 569. https://doi.org/10.1037/0021-843X.100.4.569

Nosek, B. A., & Errington, T. M. (2020). What is replication?. *PLOS Biology*, 18(3), e3000691. https://doi.org/10.1371/journal.pbio.3000691

Nosek, B. A., Hardwicke, T. E., Moshontz, H., Allard, A., Corker, K. S., Dreber, A., ... & Vazire, S. (2022). Replicability, robustness, and reproducibility in psychological science. *Annual Review of Psychology*, 73(1), 719-748.

Palan, S., & Schitter, C. (2018). Prolific. ac—A subject pool for online experiments. *Journal of Behavioral and Experimental Finance*, *17*, 22-27. https://doi.org/10.1016/j.jbef.2017.12.004

Radloff, L. S. (1977). The CES-D scale: A self-report depression scale for research in the general population. *Applied psychological measurement*, *1*(3), 385-401. https://doi.org/10.1177/014662167700100306

Russell, D., Peplau, L. A., & Cutrona, C. E. (1980). The revised UCLA Loneliness Scale:

concurrent and discriminant validity evidence. *Journal of personality and social*

psychology, *39*(3), 472. https://doi.org/10.1037/0022-3514.39.3.472

Sargent, R. H., & Newman, L. S. (2021). Pluralistic ignorance research in psychology: A scoping

review of topic and method variation and directions for future research. *Review of General*

Psychology, *25*(2), 163-184. https://doi.org/10.1177/1089268021995168

Schanck, R. L. (1932). A study of a community and its groups and institutions conceived of as

behaviors of individuals. *Psychological Monographs*, *43*(2).

https://doi.org/10.1037/h0093296

Sharpe, N. R., & Roberts, R. A. (1997). The relationship among sums of squares, correlation

coefficients, and suppression. *The American Statistician*, *51*(1), 46-48.

Sokoloski, R., Markowitz, E. M., & Bidwell, D. (2018). Public estimates of support for offshore

wind energy: False consensus, pluralistic ignorance, and partisan effects. *Energy Policy*,

112, 45-55. https://doi.org/10.1016/j.enpol.2017.10.005

Steers, M. L. N., Wickham, R. E., & Acitelli, L. K. (2014). Seeing everyone else's highlight reels:

How Facebook usage is linked to depressive symptoms. *Journal of Social and Clinical*

Psychology, *33*(8), 701-731. https://doi.org/10.1521/jscp.2014.33.8.701

Tandoc, E. C., Ferrucci, P., & Duffy, M. (2015). Facebook use, envy, and depression among

college students: Is Facebooking depressing? *Computers in human behavior*, *43*, 139-146.

https://doi.org/10.1016/j.chb.2014.10.053

Thompson, F. T., & Levine, D. U. (1997). Examples of easily explainable suppressor variables in

multiple regression research. *Multiple Linear Regression Viewpoints*, *24*(1), 11-13.

Treynor, W., Gonzalez, R., & Nolen-Hoeksema, S. (2003). Rumination reconsidered: A

psychometric analysis. *Cognitive therapy and research*, *27*, 247-259.

https://doi.org/10.1023/A:1023910315561

Verduyn, P., Lee, D. S., Park, J., Shablack, H., Orvell, A., Bayer, J., Ybarra, O., Jonides, J., &

Kross, E. (2015). Passive Facebook usage undermines affective well-being: Experimental

and longitudinal evidence. *Journal of experimental psychology. General*, *144*(2), 480–

488. https://doi.org/10.1037/xge0000057

Vogel, E. A., Rose, J. P., Roberts, L. R., & Eckles, K. (2014). Social comparison, social media,

and self-esteem. *Psychology of popular media culture*, *3*(4), 206.

Vogel, E. A., Rose, J. P., Okdie, B. M., Eckles, K., & Franz, B. (2015). Who compares and

despairs? The effect of social comparison orientation on social media use and its

outcomes. *Personality and Individual Differences, 86,* 249–256.

https://doi.org/10.1016/j.paid.2015.06.026

Wang, J. L., Wang, H. Z., Gaskin, J., & Hawk, S. (2017). The mediating roles of upward social

comparison and self-esteem and the moderating role of social comparison orientation in

the association between social networking site usage and subjective well-being. *Frontiers

in psychology, 8*, 771. https://doi.org/10.3389/fpsyg.2017.00771

Whillans, A. V., Christie, C. D., Cheung, S., Jordan, A. H., & Chen, F. S. (2017). From

Misperception to Social Connection: Correlates and Consequences of Overestimating

Others' Social Connectedness. *Personality and Social Psychology Bulletin*, 43(12), 1696–

1711. https://doi.org/10.1177/0146167217727496

Yonelinas, A. P., & Ritchey, M. (2015). The slow forgetting of emotional episodic memories: An

emotional binding account. *Trends in cognitive sciences*, *19*(5), 259-267.

https://doi.org/10.1016/j.tics.2015.02.009

Young, S. D., & Jordan, A. H. (2013). The influence of social networking photos on social norms

and sexual health behaviors. *Cyberpsychology, behavior, and social networking*, *16*(4),

243-247. https://doi.org/10.1089/cyber.2012.0080

Yu, E., & Kim, H. (2020). Is she really happy? A dual-path model of narcissistic self-presentation

outcomes for female Facebook users. *Computers in Human Behavior*, *108*, 106328.

https://doi.org/10.1016/j.chb.2020.106328

Zwaan, R. A., Etz, A., Lucas, R. E., & Donnellan, M. B. (2018). Making replication mainstream.

Behavioral and Brain Sciences, *41*, e120.

The mis-estimation of others' emotions:
Replication of Jordan et al. (2011) emotional prevalence pluralistic ignorance with extensions examining emotional intensity and social comparison orientation

Supplementary

Contents

Analysis of the **target article**

Target article methods

We summarized the hypotheses from the target article in Table S1.

Table S1

Jordan et al. (2011): Summary of hypotheses

Study	Hypothesis	Description of hypotheses in the present replication
1b	2a	People underestimate the prevalence and extent of others' negative emotional experiences.
	2b	People do not underestimate the prevalence and extent of others' positive emotional experiences.
		[Reframed from null hypothesis: Underestimation errors would be stronger for negative experiences than for positive experiences.]
3	4	Stronger misperception of their peers' emotional lives is associated with lower well-being: greater depressive symptoms, loneliness, rumination, lower life satisfaction, and subjective happiness.

Independent variables (IV)

The degree of pluralistic ignorance was the independent variable across Study 1b and 3 in the target article. Operationalization of the constructs required estimation of emotional events with different valence (positive versus negative) and on different perspectives (self versus others).

Dependent variables (DV)

Five well-being measures, namely loneliness, brooding, depression, life satisfaction, and subjective happiness, were the dependent variables in the original text. Table S2 summarizes the description, questions, and the reported reliability of all dependent variables.

Table S2

Summary of well-being measures and their reported reliability

Well-being measures	Description of the scale	Cronbach's α
Loneliness	Participants were asked to complete the 8-item short-form version of the UCLA Loneliness Scale (Hays & DiMatteo, 1987; Russell, Peplau, & Cutrona, 1980). Items include "There is no one I can turn to." and "I can find companionship when I want it." (1 = *Never* to 4 = *Often*)	.81
Rumination/brooding	Participants were asked to complete the Brooding subscale (5-item) of the Ruminative Responses to Depression Questionnaire (Nolen-Hoeksema, 1991; Treynor, Gonzales & Nolen-Hoeksema, 2003). Items include "Think "What am I doing to deserve this?" and "Think about a recent situation, wishing it had gone better." (1 = *Almost never* to 4 = *Almost always*)	.72
Depression	Participants were asked to complete the 10-item short-form version of the Center for Epidemiologic Studies Depression Scale (Cole, Rabin, Smith, & Kaufman, 2004; Radloff, 1977). Items include "I was bothered by things that usually don't bother me." and "I had trouble keeping my mind on what I was doing."	.73
Satisfaction with life	Participants were asked to complete the 5-item Satisfaction with Life Scale (SWLS) (Diener, Emmons, Larsen, & Griffin, 1985).	.84
Subjective Happiness	Participants were asked to complete The 4-item Subjective Happiness Scale (SHS) (Lyubomirsky & Lepper, 1999).	.80

Target article results

Sample size before and after exclusions

The sample sizes for Studies 1b and 3 were 80 and 104 respectively without any exclusions mentioned.

Included sample description

Both studies reported in the target article yielded their samples among students. Study 1b recruited first-year university students (45 female and 35 male) from a medium-sized West Coast university who completed a paper-and-pen questionnaire format. Study 3 collected data using an internet survey. Participants (53 female and 51 male) were recruited from a pool of online study participants from a medium-sized West Coast university, without specifying if it is the same university in Study 1b.

One sample experiment [no manipulation experiments]

Study 1b

For each emotional experience, a one-sample t-test was conducted to compare the estimated prevalence of the experiences for their peers and the reported prevalence reflected by their self-estimation. The estimated, reported prevalence and results of the one-sample t-test were taken from page 126 (Study 1b) from the target article and presented below.

Table 2. Estimation Errors for Negative and Positive Emotional Experiences (Study 1b)

	Percentage of classmates who have had experience in past 2 weeks			
Experience	Estimated	SD	Reported	Estimation error (estimated − reported)
Negative experiences				
Had fight/argument	29.5	18.6	42.5	−13.0***
Thought about distant friends/family	53.7	25.1	82.5	−28.8***
Thought about enormous workload	75.3	20.8	87.5	−12.2***
Was rejected by boy/girl	22.4	12.6	31.3	−8.9***
Received low grade	44.1	19.6	60.0	−15.9***
Thought about bad personal health habits	45.9	24.3	70.0	−24.1***
Positive experiences				
Received high grade	62.0	19.6	65.0	−3.0
Attended fun party	62.2	17.6	41.3	+20.9***
Participated in athletics	53.7	20.5	40.0	+13.7***
Went out with friends	57.6	20.3	45.0	+12.6***
Talked to distant friends/family	61.7	18.8	70.0	−8.3***
Had great meal	61.5	24.1	63.8	−2.3

***$p < .001$.

Participants underestimated the prevalence of all six negative emotional experiences and one positive emotional experience. They overestimated three positive emotional experiences and the estimated prevalence of the other positive emotional experiences did not differ from the reported prevalence.

The degree of freedom, p-values, and confidence intervals of the one-sample t-tests was not reported.

In study 1b, participants underestimated 17.2% on average for the negative experiences with a standard deviation of 7.7%, which differed from that for positive experiences, which was an overestimation of 5.6% with a standard deviation of 11.7%. An independent t-test was conducted with a t-statistics value equal to 3.99; the degree of freedom was 10, the p-value was smaller than 0.01 and the effect size expressed in Cohen's d equaled 2.52.

Overall, one sample t-test was conducted and found that the estimation error for negative experiences differed from zero. T-statistics equaled 5.47; degrees of freedom were 5, and the p-value was smaller than 0.01, but the effect size expressed in Cohen's d was not reported. On the contrary, the estimation

error for positive experiences did not differ from zero, with a t-statistics value equal to 1.18; the degree of freedom was 5, and the p-value was not reported.

Study 3

Participants estimated the prevalence of others' emotional experiences repeatedly in Study 3 and see Table 4 for the results of the one-sample t-test. The screenshot was taken from page 130 from the target article.

Table 4. Estimation Errors for the Prevalence of Negative and Positive Emotional Experiences (Study 3)

Experience	Percentage of classmates who have had experience in past 2 weeks			Estimation error (estimated – reported)
	Estimated	SD	Reported	
Negative experiences				
Had fight/argument	29.5	17.7	43.3	−13.8***
Thought about distant friends/family	59.3	23.4	85.6	−26.3***
Thought about enormous workload	78.1	15.6	89.4	−11.3***
Was rejected by boy/girl	19.1	11.3	37.5	−18.4***
Received low grade	37.3	16.1	60.6	−23.3***
Thought about bad personal health habits	44.8	23.5	79.8	−35.0***
Positive experiences				
Received high grade	64.1	22.1	64.4	−0.3
Attended fun party	54.5	20.6	41.3	+13.2***
Participated in athletics	56.6	19.7	49.0	+7.6***
Went out with friends	49.0	21.4	37.5	+11.5***
Talked to distant friends/family	52.6	20.2	62.5	−9.9***
Had great meal	54.3	20.9	53.8	+0.5

***$p < .001$.

Participants underestimated the prevalence of all six negative emotional experiences and one positive emotional experience. They overestimated two positive emotional experiences and the estimated prevalence of the other two positive emotional experiences did not differ from the reported prevalence.

The degree of freedom, p-values, and confidence intervals of the one-sample t-tests was not reported.

In study 1b, participants underestimated 21.4% on average for the negative experiences with a standard deviation of 8.7%, which differed from that for positive experiences, which was an overestimation of 3.8% with a standard deviation of 8.7%. An independent t-test was conducted with a t-statistics value equal to 4.99; the degree of freedom was 10, the p-value was equaled 0.001 and the effect size expressed in Cohen's d equaled 3.16.

Overall, one sample t-test was conducted and found estimation error for negative experiences. T-statistics equaled 5.99; degrees of freedom were 5, and the p-value was smaller than 0.01, but the effect size expressed in Cohen's d was not reported. On the contrary, the estimation error for positive experiences did not differ from zero, with a t-statistics value equal to 1.06; the degree of freedom was 5, and the p-value was not reported.

Correlational study

See Table 5 for the descriptives of the well-being measures taken from page 130 of the target article.

Table 5. Predicting Well-Being Measures From Prevalence Estimates (Study 3)

Life-quality measure	M	SD	Cronbach's α	β for negative prevalence estimate	β for positive prevalence estimate	R^2
Negative						
Loneliness	17.0	4.5	.81	−.30**	.08	.08
Rumination/brooding	10.7	3.3	.72	−.28**	.17	.08
Depressive symptoms	8.8	4.7	.73	.00	−.02	.00
Positive						
Satisfaction with life	24.5	4.6	.84	.23*	−.37***	.15
Subjective happiness	4.8	1.1	.80	.19	−.16	.05

*$p < .05$. **$p < .01$. ***$p < .001$.

Authors from the target article computed the average of each participant's prevalence estimates for the six negative experiences ("negative prevalence estimate") had a mean of 44.7 and a standard deviation of 10.1. Similarly, the "positive prevalence estimate" had a mean of 55.2 and a standard deviation of 9.2.

Positive prevalence estimates positively correlated with negative prevalence estimate with a Pearson Correlation equaled .27 and a p-value smaller than .01.

The authors built regression models to examine the predictive abilities of positive and negative prevalence estimates on well-being. Negative prevalence estimates negatively related to loneliness with beta equaled -.30, t-statistics equaled 3.02, degree of freedom equaled 101, and p-value smaller than .01. Similarly, negative prevalence estimates negatively related to brooding with beta equaled -.28, t-statistics equaled 2.86, degree of freedom equaled 101, and p-value smaller than .01. In addition, negative prevalence estimates positively related to the level of life satisfaction with beta equaled -.23, t-statistics equaled 2.44, degree of freedom equaled 101, and p-value smaller than .02. Consistent with their predictions, positive prevalence estimates negatively related to the level of satisfaction with beta equaled -.37, t-statistics equaled 3.89, degree of freedom equaled 101, and p-value smaller than .001.

Effect size calculations of the original study effects

Please see files:

        Jordan_etal_2011_rep_ext-power-analyses.RMD/html

for Rmarkdown output of calculated effect sizes of the target article with power analyses.

The following are the effect size from R^2 reported in the target article converted to f^2

Effect size f^2 of the prediction of loneliness from prevalence estimates
= 0.08 / (1- 0.08)
= 0.08695652

Effect size f^2 of the prediction of brooding from prevalence estimates
= 0.08 / (1- 0.08)
= 0.08695652

Effect size f^2 of the prediction of life satisfaction from prevalence estimates
= 0.15 / (1- 0.15)
= 0.17647058823

Effect size f^2 of the prediction of subjective happiness from prevalence estimates
= 0.05 / (1- 0.05)
= 0.05263157894

Table S3

*Effects the original study effects and the required sample size of Studies 1b and 3*

Study	Statistical Test	Test Objective	Original result [negative emotions, positive emotions]	Calculated/ Reported Cohen's d, 95%CI	Target article's sample size	Required sample size
1b	Independent sample t-test	Comparing average estimation error between positive and negative prevalence estimates	t(10) = 3.99*	1.17/ 2.52 [-0.09, 2.39]	80	34
1b	One-sample t-test	Compare if the average estimation error for negative zero differed from zero	t(5) = 5.47**	2.45 [0.58, 4.27]	80	4
1b	One-sample t-test	Compare if the average estimation error for positive zero differed from zero	t(5) = 1.18	0.53 [-0.44, 1.44]	80	40
3	Independent sample t-test	Comparing average estimation error between positive and negative prevalence estimates	t(10) = 4.99**	2.02/ 3.16 [0.56, 3.42]	104	14
3	One-sample t-test	Compare if the average estimation error for negative zero differed from zero	t(5) = 5.99*	2.68 [0.68, 4.65]	104	4
3	One-sample t-test	Compare if the average estimation error for positive zero differed from zero	t(5) = 1.06,	0.47 [-0.48, 1.38]	104	51
3	Linear regression	Examine how prevalence estimates (negative; positive) predicted loneliness	β = -0.30**, 0.08 t(101) = 3.02**	0.08 [-0.02, 0.18]	104	181
3	Linear regression	Examine how prevalence estimates (negative; positive) predicted brooding	β = -0.28**, 0.17 t(101) = 2.86**	0.08 [-0.02, 0.18]	104	181
3	Linear regression	Examine how prevalence estimates (negative; positive) predicted depressive symptoms	β = 0.00, -0.02	0.00 [0.00, 0.00]	104	/
3	Linear regression	Examine how prevalence estimates (negative; positive) predicted life satisfaction	β = 0.23*, -0.37*** t(101) = 2.44*, 3.89***	0.15 [0.29, 0.27]	104	91
3	Linear regression	Examine how prevalence estimates (negative; positive) predicted subjective happiness	β = 0.19, -0.16	0.05 [-0.03, 0.13]	104	297

Note. The column reporting original results of the target article were extracted from page 125, 129 - 130 in Jordan et al. (2011). *$p < .05$.**$p < .01$. ***$p < .001$. Power analyses aiming for 95% power at alpha = .05,

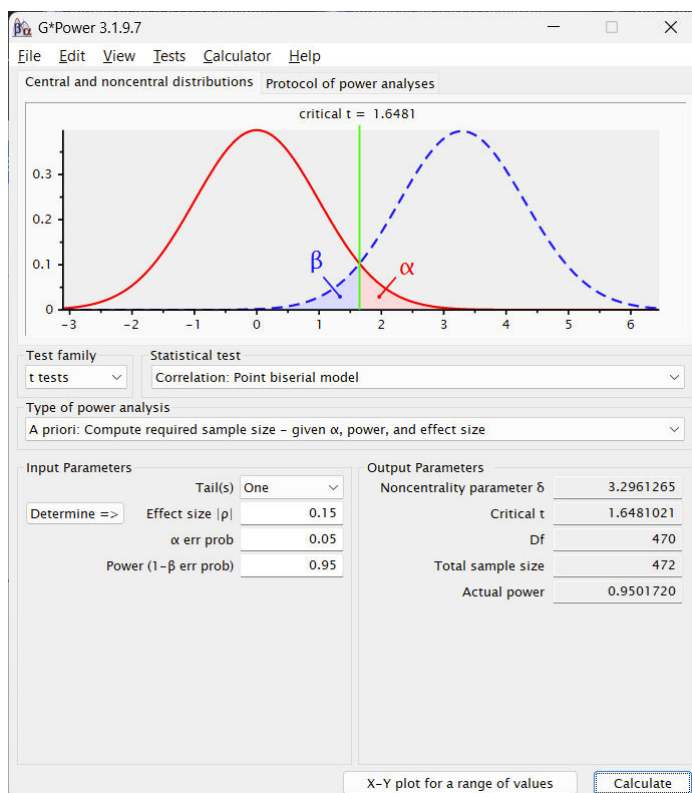## Power analyses

Screenshots from G*Power (Faul et al., 2017).

Aiming for paired/one-sample t-test with SESOI *d* = 0.2
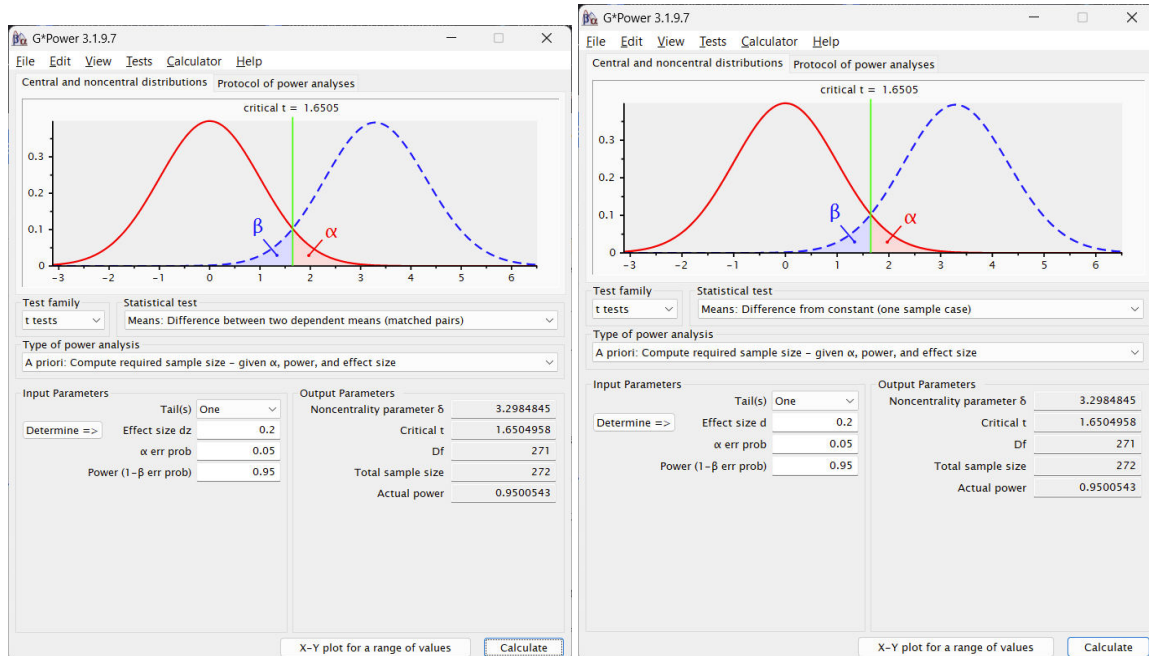
Aiming for correlations of SESOI *r* = .15

Handling outliers: Strategy

In this study, we pre-registered that all the data would be included in our analyses, and that we would not identify outliers.

Materials and scales used in the **replication + extension experiment**

Please see the Qualtrics export files in the OSF for the clearest most up-to-date files, including the full survey. Below is provided as a brief summary only, and in any discrepancies the Qualtrics is the one to consult and use.

We used Qualtrics used for data collection of Studies 1b and Study 3 on the online platform CloudResearch Connect.

First, participants read the consent form and study outline. Thereafter, they read the statements concerning bonuses for prediction accuracy.

Participants in our studies then went into all four conditions and they estimated the prevalence, on a scale from 0-100, of the different emotional experiences. Before indicating the required prevalence, we added comprehension checks to make sure they paid attention and understood the rating instructions:

Whose emotional experiences are you rating?
      o My own emotional experiences
      o Other US American student participants on Prolific like me
      o My friends and family

In this task, what are you rating in regards to these emotional experiences?
      o Prevalence
      o Intensity
      o Both prevalence and intensity

The instructions across the four conditions were outlined below.

OP_Intro Estimation of others' emotional experiences: Prevalence

In this task, you are asked to estimate the prevalence of emotional experiences.
Please estimate the percentage of other US American student participants on Prolific taking the survey like you who had had, sometime in the past 2 weeks, each of the following emotional experiences.

OI_Intro Emotional Intensity of emotional experience for others

In this task, you are asked to estimate the intensity of the emotional experiences of those who have experienced this emotion in the past two weeks.
Please try and estimate the emotional intensity for other US American student participants on Prolific taking the survey like you who have experienced this emotion (1 = lowest emotional intensity, and 100 = highest emotional intensity.)

S_intro Your own emotional experiences: Prevalence and Intensity

In this task, you are asked to indicate a list of emotional experiences, whether you have experienced those sometime in the past 2 weeks, and if you have - the intensity of the emotion.

0 means: You have not experienced this emotion in the past two weeks.

1-100 means: You have experienced this emotion at least once in the past two weeks.

1 = lowest emotional intensity, and 100 = highest emotional intensity.

Upon giving instructions, participants were presented with the following twelve emotional experiences in randomized order.

1. Had fight/argument
2. Thought about distant friends/ family
3. Thought about enormous workload
4. Was rejected by boy/girl
5. Received low grade
6. Thought about bad personal health habits
7. Received high grade
8. Attended fun party
9. Participated in athletics
10. Went out with friends
11. Talked to distant friends/ family
12. Had great meal

To avoid potential confounding effects due to differential perception in test items, clarification questions were asked and participants could indicate their perceptions for up to 100 words.

clarify_intro **Perceptions of the test items**

The items we used in this survey are based on classic research, and so we are using items that other research previously used. To ensure we are aligned with your view on those items, we would like to inquire about your understanding of some of these items.

Therefore, on this page, you are asked to share your understanding of the items you rated:

clarify_grade What is a "bad" grade? How do you classify a grade as being a bad grade? Can you give a quick example? (1-2 sentences)

clarify_workload What is an "enormous" workload? How do you classify workload as being an enormous workload? Can you give a quick example? (1-2 sentences)

clarify_health What is a "bad" personal health habit? How do you classify personal health habit as being bad? Can you give a quick example? (1-2 sentences)

Dependent variables

Participants responded to measures of well-being in randomized order.

1. Loneliness

   Others and companionship

   In this task, you are asked to indicate how often you feel the way described in each of the following statements.

   Loneliness_rate Please read each statement carefully and indicate how often each of the statements below is descriptive of you.

   - I lack companionship.
   - There is no one I can turn to.

- I am an outgoing person.
- I feel left out.
- I feel isolated from others.
- I can find companionship when I want it.
- I am unhappy being so withdrawn.
- People are around me but not with me.

2. Brooding

Thoughts

In this task, you are asked to indicate how often you think or do as described in each of the following statements.

Brooding_rate Please read each statement carefully and indicate how often you think or do as described in each of the following statements.

- Think "What am I doing to deserve this?".
- Think "What do I always react this way?".
- Think about a recent situation, wishing it had gone better.
- Think "Why do I have problems other people don't have?".
- Think "Why can't I handle things better?".

3. Depressive symptoms

Feelings and reflections

In this task, you are asked to indicate how often you think during the past week as described in each of the following statements.

Depression_rate Please read each statement carefully and indicate how often you think during the past week as described in each of the following statements.

- I was bothered by things that usually don't bother me.
- I felt that I could not shake off the blues even with the help from my friends or family.
- I felt that I was just as good as other people.
- I had trouble keeping my mind on what I was doing.
- I felt that everything I did was an effort.
- I felt hopeful about the future.
- I felt my life had been a failure.
- I felt fearful.
- I felt lonely.
- People were unfriendly.

4. Life satisfaction

Life Satisfaction

In this task, you are asked to indicate your agreement with each of the following statements.

lifesat_rate Please read each statement carefully and indicate your agreement with each of the following statements.

- In most ways my life is close to my ideal.
- The conditions of my life are excellent.
- I am satisfied with my life.
- So far I have gotten the important things I want in life.
- If I could live my life over, I would change almost nothing.

5. Subjective happiness

Happiness

In this task, you are asked to indicate the point that you feel is most appropriate in describing you in each of the following statements.

Happy_1 In general, I consider myself…

    o 1 Not a very happy person
    o 2
    o 3
    o 4
    o 5
    o 6
    o 7 A very happy person

Happy_2 Compared to most of my peers, I consider myself:

    o 1 Less happy
    o 2
    o 3
    o 4
    o 5
    o 6
    o 7 More happy

Happy_3 Some people are generally very happy. They enjoy life regardless of what is going on, getting the most out of everything. To what extent does this characterization describe you?

    o 1 Not at all
    o 2
    o 3
    o 4
    o 5
    o 6
    o 7 A great deal

Happy_4 Some people are generally very not happy. Although they are not depressed, they never seem as happy as they might be. To what extend does this characterization describe you?

    o 1 Not at all
    o 2
    o 3

           o 4
           o 5
           o 6
           o 7 A great deal

Confidants Please indicate a rough number of friends you have around you who you feel comfortable talking to about personal emotional experiences in your life.

Participants responded to measure of social comparison orientation for our extension.

Others and I

We would like to find out how often you compare yourself with other people. In this task, you are asked to indicate how much you agree with each statement below.

sco-scale Please indicate how much you agree with each statement below.

- I often compare myself with others with respect to what I have accomplished in life.
- If I want to learn more about something I try to find out what others think about it.
- I always pay a lot of attention to how I do things compared with how others do things.
- I often compare how my loved ones (boy or girlfriend, family members, etc.) are doing with how others are doing.
- I always like to know what others in a similar situation would do.
- I am not the type of person who compares often with others.
- If I want to find out how well I've done something, I compare what I have done with how others have done.
- I often try to find out what others think who face similar problems as I face.
- I often like to talk with others about mutual opinions and experiences.
- I never consider my situation in life relative to that of other people.
- I often compare how I am doing socially (e.g. social skills, popularity) with other people.

In the subsequent part, all participants completed the three funneling questions:
- *What do you think the purpose of the last part was?*
- *Have you ever seen the materials used in this study or similar before? If yes - please indicate where*
- *Did you spot any errors? Anything missing or wrong? Something we should pay attention to in next runs? (Briefly, up to one sentence, write "none" if not relevant)*

Finally, participants were asked to fill in demographics and were debriefed. No filler items were included.

Exclusion criteria

We will run our analyses on the full sample of all those who completed the study successfully and answered all questions. Those who dropped out will not be included.

In the case of a failed replication, as a supplementary analysis and to examine any potential issues, we will also determine further findings reports with exclusions. In any case, we will report exclusions in detail with results for the full sample and results following exclusions (in either the manuscript or the supplementary.

Criteria:

1. Participants indicating a low proficiency of English (self-report < 5, on a 1-7 scale).

2. Participants who self-report not being serious about filling in the survey (self-report < 4, on a 1-5 scale).

3. Participants who indicated issues or having seen these materials before in the funneling section (manually coded).

[Note: After consent and before starting the questionnaire, participants answered a series of verification questions, which include confirming whether they are over the age of 18, stating that they are willing to pay close attention and that they understand the outline, and confirming that they were native American citizens born and raised in the United States. Failing to give a positive answer to any of the above questions will result in them seeing a message that says they are not qualified for the questionnaire. For those who were qualified for the questionnaire, they will answer a funneling section after finishing all questions, where they will be asked, on a scale from 1 to 5, how serious they were when filling out the questionnaire. Those who answered 1 and 2 will be considered outliers as they were not paying attention during the questionnaire, and their answers will not give useful insights into the phenomenon.]

Comparisons and deviations

Pre-registration plan versus final report

We have no deviations to report.

Components in your preregistration (e.g., stopping rule, analyses, hypotheses, exclusion rules)	Location of 1) preregistered decision/plan and 2) rational for decision/plan [Location / link]	Were there deviations? What type? [no / minor / major]*	If yes - describe details of deviation(s) [brief description / location / link]	Rationale for deviation [brief description / location / link]	How might the results be different if you had/had not deviated [brief description / location / link]	Date/time of decision for deviation + stage	Any additional notes
Study design	"Method" section - "Design and Procedure" of the main manuscript	no	/	/	/	/	/
Measured variables	"Method" section - "Measures" of the main manuscript	no	/	/	/	/	/
Exclusion criteria	"Exclusion criteria" section in the supplementary	no	/	/	/	/	/
IV	"Method" section main manuscript	no	/	/	/	/	/
DV	"Method" section main manuscript	no	/	/	/	/	/
Data analysis	"Method" and "Results" section of the main manuscript	no	/	/	/	/	/

notes: *Categories for deviations: Minor - Change probably did not affect results or interpretations; Major - Change likely affected results or interpretation.

Additional information about the study

Data collection procedures:

This study was conducted on Prolific with American participants. We imposed the following settings in recruiting our participants:

1. We first pretested the survey duration with 30 participants with £1.8 to make sure our time run estimate was accurate. As the survey duration was longer than expected, we awarded our pre-test participants with a £0.2 as a pay adjustment. The data of the 30 participants were not analyzed other than to assess survey completion duration and needed pay adjustments. The pretest participants' responses were included in the final analysis.
2. For the full-run data collection, we initially awarded our participants ($N = 594$) £2.0 and eventually increased it to £2.3 due to the longer-than-expected survey duration.
3. The expected completion time was set at 12 minutes in advance.
4. The most time we allowed each worker to complete the study was 30 minutes.
5. We limited all workers' HIT Approval Rate to be between 95% and 100%.
6. We limited each worker's number of HITs approved to be between 5,000 and 100,000.
7. We blocked Suspicious Geocode Locations and Universal Exclude List Workers.
8. We blocked duplicate IP addresses and duplicate geolocation.
9. We restricted workers' location to be in the U.S and criteria to students.
10. We allowed an answer time of 30 minutes.
11. We did not restrict answering from specific devices.
12. The top 10 participants who made the most accurate predictions received a 10 USD bonus as a reward. If there were more than 10 participants with the same top prediction score, 10 participants who were randomly selected were rewarded.

Additional Tables and Figures

Social comparison orientation associations

[We note the following were weak effects and in a series of many tests for interactions. We therefore caution against over-interpretation, and report these here as exploratory. We recommend focusing on the much stronger and clearer main effects reported in the main manuscript.]

We conducted a moderation analysis and found that social comparison orientation only moderated the relations between negative prevalence estimates and subjective happiness, $\beta = .00$, $z(591) = 2.22$, $p = .03$. In particular, we found that participants with lower-than-average social comparison orientation showed a stronger negative association between negative emotions misestimation and happiness, $\beta = -.05$, $z(591) = -2.54$, $p = .01$.

Social comparison orientation moderated the relations between positive prevalence estimates and life satisfaction, $\beta = .01$, $z(591) = 2.00$, $p = .05$. In particular, we found that participants with higher-than-average social comparison orientation showed a stronger positive relationship between positive prevalence estimates and life satisfaction, $\beta = .10$, $z(591) = 3.83$, $p < .001$.

Table S4

Interaction [Extension]: Social comparison orientation and misestimation in predicting well-being

Variable	β of the interaction between negative prevalence estimate and well-being	p	β of the interaction effect between positive prevalence estimate and well-being indicator	p
Interaction: Social comparison orientation				
Loneliness	-0.00 [-0.01, 0.00]	.552	-0.00 [-0.01, 0.00]	.681
Brooding	0.00 [-0.00, 0.01]	.075	-0.00 [-0.00, 0.00]	.827
Depression	0.00 [-0.00, 0.01]	.817	-0.00 [-0.01, 0.00]	.811
Life satisfaction	0.01 [-0.00, 0.01]	.087	0.01 [0.00, 0.01]	.046*
Subjective happiness	0.00 [0.00, 0.01]	.027*	0.00 [-0.00, 0.01]	.082

Note. Linear regression, N = 500. CI = 95% confidence intervals. *$p < .05$; **$p < .01$; ***$p < .001$.

Figure S1

Social comparison orientation: Plot of interaction with misestimation of negative emotions in predicting subjective happiness
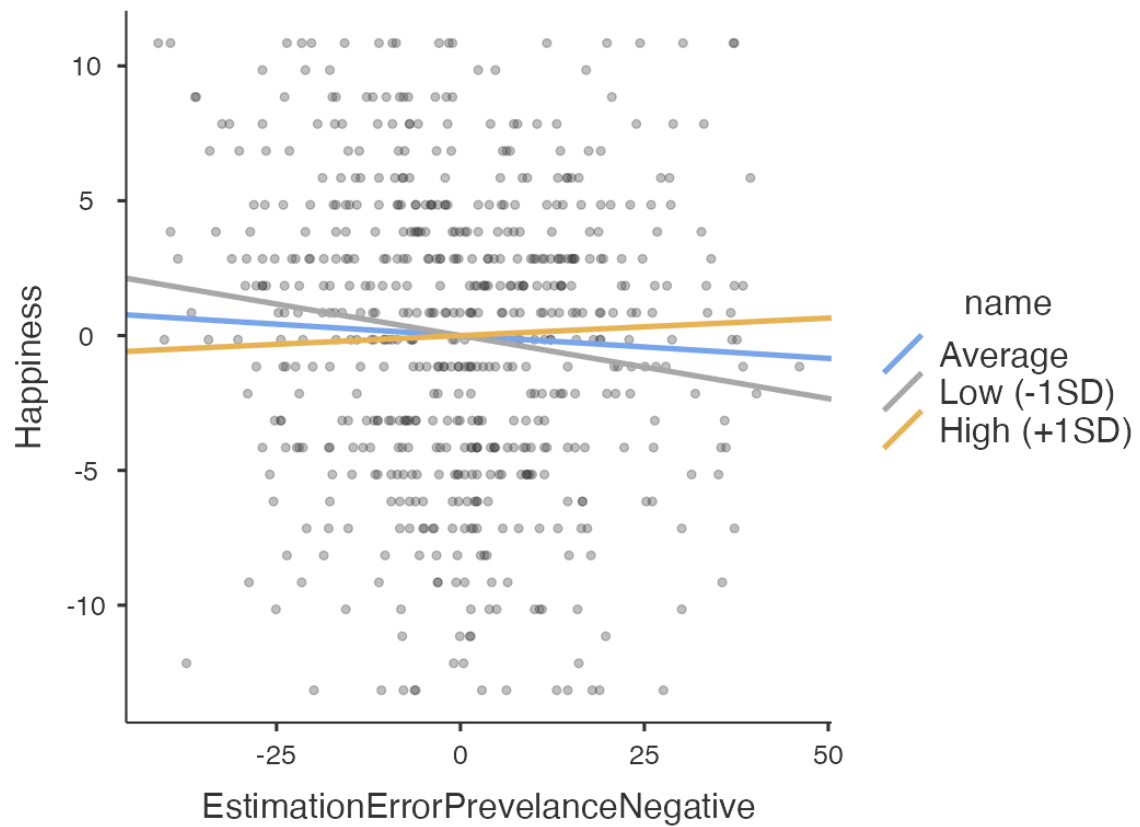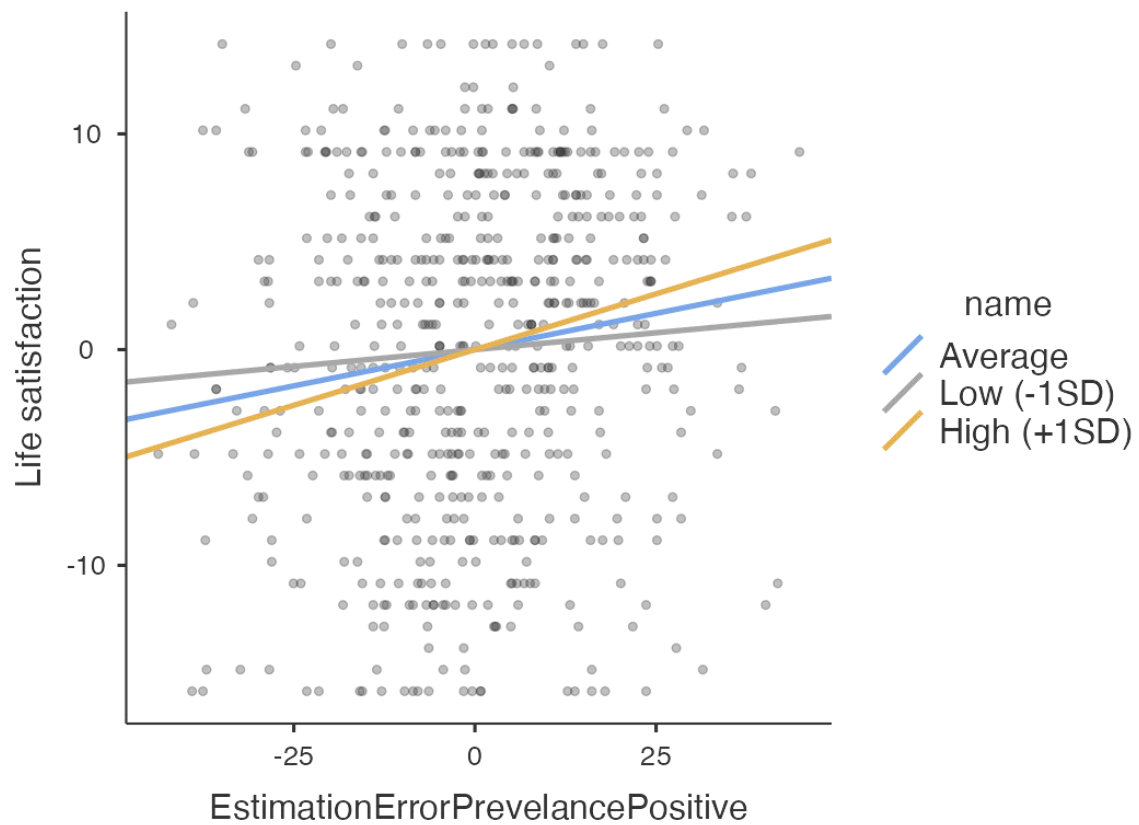
Figure 2

Social comparison orientation: Plot of interaction with misestimation of positive emotions in

predicting life satisfaction

Replication evaluation

Replication closeness

Lebel, McCarthy, Earp, Elson, and Vanpaemel (2018):

Target similarity	Highly similar				Highly dissimilar
Category	Direct replication			Conceptual replication	
Design facet	**Exact replication**	**Very close replication**	**Close replication**	**Far replication**	**Very far replication**
Effect/hypothesis	Same/similar	Same/similar	Same/similar	Same/similar	Same/similar
IV construct	Same/similar	Same/similar	Same/similar	Same/similar	Different
DV construct	Same/similar	Same/similar	Same/similar	Same/similar	Different
IV operationalization	Same/similar	Same/similar	Same/similar	Different	
DV operationalization	Same/similar	Same/similar	Same/similar	Different	
Population (e.g. age)	Same/similar	Same/similar	Same/similar	Different	
IV stimuli	Same/similar	Same/similar	Different		
DV stimuli	Same/similar	Same/similar	Different		
Procedural details	Same/similar	Different			
Physical setting	Same/similar	Different			
Contextual variables	Different				

Figure 3. Criteria for evaluation of replications by LeBel et al. (2018).
A classification of relative methodological similarity of a replication study to an original study. "Same" ("different") indicates the design facet in question is the same (different) compared to an original study. IV = independent variable. DV = dependent variable. "Everything controllable" indicates design facets over which a researcher has control. Procedural details involve minor experimental particulars (e.g., task instruction wording, font, font size, etc.).
"Similar" category was added to the Lebel et al. (2018) typology to refer to minor deviations or extensions aimed to adjust the study to the target sample that are not expected to have major implications on replication success. See Olsson-Collentine, van Assen, and Wicherts (2020) on meta-analysis showing minor to no expected impact due to variations in sample population or setting.

Replication versus the original

LeBel, Vanpaemel, Cheung, and Campbell (2019) criteria:
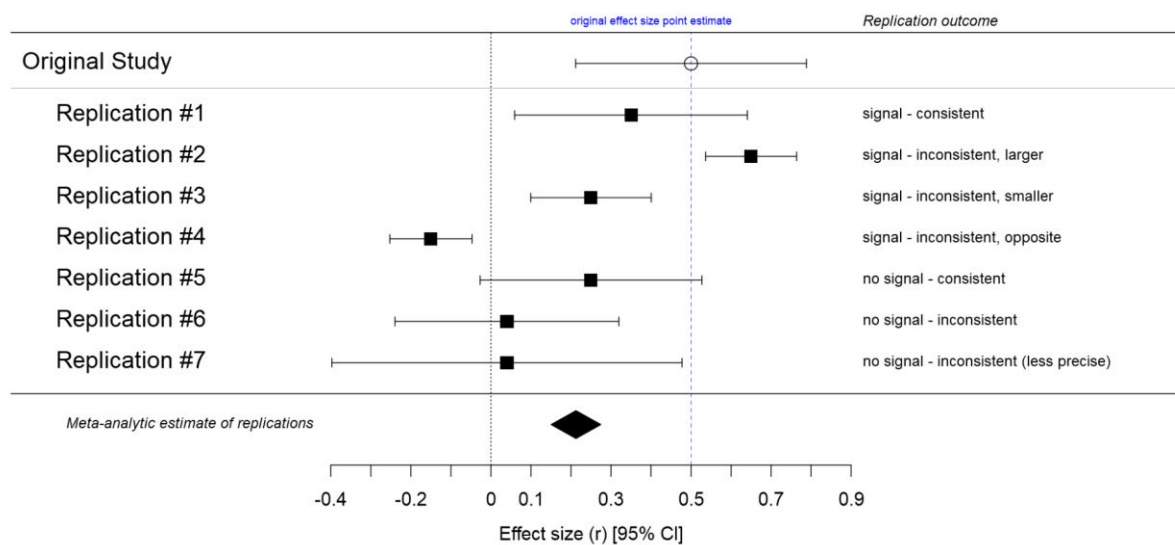

Figure S2. Interpretation criteria for evaluation of replications outcomes by LeBel et al. (2019), if the original study detected a signal. A simplified replication taxonomy for comparing replication effects confidence intervals to target article original effect sizes.
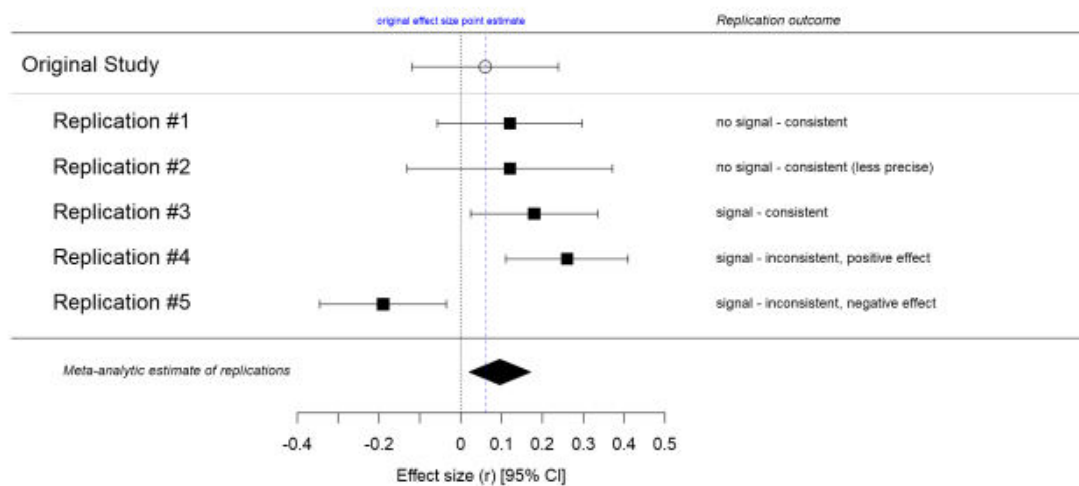

Figure S3. Interpretation criteria for evaluation of replications outcomes by (LeBel et al., 2019), if the original study failed to detect a signal.