ORIGINAL RESEARCH REPORT

The Mirror Effect: A Preregistered Replication

Jean Monéger, Armand Chatard and Leila Selimbegović

When individuals are exposed to their own image in a mirror, known to increase self-awareness, they may show increased accessibility of suicide-related words (a phenomenon labeled "the mirror effect"; Selimbegović & Chatard, 2013). We attempted to replicate this effect in a pre-registered study (N = 150). As in the original study, self-awareness was manipulated using a mirror and recognition latencies for accurately detecting suicide-related words, negative words, and neutral words in a lexical decision task were assessed. We found no evidence of the mirror effect in pre-registered analyses. A multiverse analysis revealed a significant mirror effect only when excluding extreme observations. An equivalence TOST test did not yield evidence for or against the mirror effect. Overall, the results suggest that the original effect was a false positive or that the conditions for obtaining it (in terms of statistical power and/or outlier detection method) are not yet fully understood. Implications for the mirror effect and recommendations for pre-registered replications are discussed.

Keywords: Self-awareness; Suicide thought accessibility; Median Absolute Deviation

1. Introduction

Self-awareness can be defined as the capacity to direct attention towards oneself (self-focus state) and to engage in reflexive thought about oneself (Carver & Scheier, 1981). Studies investigating the effects of self-awareness manipulate self-focus in a variety of manners: by displaying participants' names (Silvia, 2012; Silvia & Phillips, 2013), by exposing participants to their own voices (Ickes, Wicklund, & Ferris, 1973), or to their mirror-reflected images (Bender, O'Connor, & Evans, 2018; Dijksterhuis & van Knippenberg, 2000; Gendolla & Richter, 2010; Heine, Takemoto, Moskalenko, Lasaleta, & Henrich, 2008; Hutton & Baumeister, 1992). The latter might be the most common manipulation of self-awareness, as shown by experiments focusing on the effects of the presence (vs. absence) of a mirror in various domains such as implicit behavior priming (Dijksterhuis & van Knippenberg, 2000), cardiovascular effort and motivation (Gendolla & Richter, 2010; Silvia, 2012, Study 3), resistance to persuasion (Hutton & Baumeister, 1992), or semantic category activation (Selimbegović & Chatard, 2013). Research also suggests that although the mere presence of a mirror might seem like a mundane detail, it can bring about very negative consequences, such as lowering self-esteem (Heine et al., 2008; Ickes et al., 1973) and facilitating access to self-destructive thoughts

(Chatard & Selimbegović, 2011; see also Fejfar & Hoyle, 2000; Mor & Winquist, 2002; Smith & Greenberg, 1981). The present study focused on one specific consequence of self-awareness, the mirror effect (Selimbegović & Chatard, 2013).

Selimbegović and Chatard (2013) suggested that mirror exposure may facilitate the detection of suicide-related words in a lexical decision task. Importantly, the authors did not claim that self-awareness alone could make people more suicidal. Instead, the core idea was that self-awareness would activate a motivation to avoid this aversive state and could therefore bring to mind escape-related constructs. Suicide being an efficient and radical means to escape self-awareness (Baumeister, 1990), mirror exposure could inadvertently increase the accessibility of suicide-related words. The results of an experiment were consistent with this prediction (Selimbegović & Chatard, 2013): participants were faster at correctly identifying suicide-related words when tested in front of a mirror, rather than in a no-mirror control condition. This finding was consistent with previous research and theorizing showing (a) that self-awareness activates unfavorable comparison between one's actual self-representation and one's ideal self-representations (Duval & Wicklund, 1972; Scheier & Carver, 1983; Silvia & Duval, 2001), (b) that when a specific motivation is pursued the most effective means to reach that goal is activated (Eitam & Higgins, 2010; Kruglanski et al., 2002), and (c) that unfavorable comparison between the actual and ideal self can be sufficient to increase the accessibility of suicide-related thoughts (Chatard & Selimbegović, 2011; Chatard, Selimbegović, Pyszczynski, & Jaafari, 2017; Tang, Wu, & Miao, 2013).

Center for Research on Cognition and Learning (CeRCA), UMR 7295, University of Poitiers and National Center for Scientific Research (CNRS), FR

Corresponding author: Leila Selimbegović (leila.selimbegovic@univ-poitiers.fr)

1.1. Aims of the current replication

The mirror effect brings a new perspective to the comprehension of self-awareness by positing that one of the simplest and most mundane acts of self-focusing (i.e. looking at one's mirror reflection) can inadvertently lead to the activation of escape responses among normal (i.e., non clinical) populations. These theoretical and practical interests encouraged us to test the reliability of the mirror effect in an attempt to conduct a replication as close to the original as possible.

In this study, the aim was to replicate the finding that self-awareness alone facilitates access to suicide-related words measured by a lexical decision task similar to the original one. Because the mirror effect may seem surprising at first sight, we decided to preregister the hypotheses and the analysis plan at the Open Science Framework website for a maximum of transparency (https://osf.io/ek2gp/). Another main theoretical interest registered prior to data collection was to assess the possible emotional mechanisms involved in the mirror effect. In order to do that, post- experimental measures of shame and guilt were added. Since these constructs were assessed at the end of the experiment, they could not influence the replicated mirror effect. In addition, the analyses of these indicators were conditional to the detection of a significant mirror effect. As the mirror effect was not replicated in preregistered analyses, and as these indicators had no significant relation to the mirror effect, they will not be discussed further. This paper will thus focus on the replication of the mirror effect.

1.2. Power analysis

The original mirror effect was of moderate size (Cohen's $d = 0.43$, 95% CI [0.038; 0.827]). Power analysis with G*Power software (Erdfelder, Faul, & Buchner, 1996) indicated that to have 80% statistical power to detect such an effect, the required sample size was 136 participants (one-tailed tests, directional hypothesis). We therefore decided in advance to recruit 150 participants for this study to anticipate possible exclusions.

2. Method

A hundred and fifty first-year psychology students at a French University (131 women and 19 men, M_{age} = 18.83 years old) participated in the study in exchange for course credits. In accordance with the preregistration, one participant was excluded from the analyses because s/he failed to complete the selves questionnaire, and one participant was excluded because s/he was not a native French speaker. Thus, the final sample consisted of 148 students.

2.1. Material

2.1.1. Mirror manipulation

Self-awareness was manipulated by mirror exposure (29 × 29 cm, or 11.42 × 11.42 inches). Participants were randomly assigned to one of the two self-awareness conditions using a random number generator (https://www.randomizer.org/). Half of them did the entire experiment while their mirror-reflected image was visible in their peripheral field of vision, while the other half of the participants were assigned to the control condition in which the mirror was facing the wall.

2.1.2. The Selves questionnaire

In the original study, self-discrepancy salience was manipulated orthogonally to mirror exposure, in order to test the moderating effect of this variable. Although self-discrepancy salience did not qualify the mirror exposure effect in the original study, we kept this manipulation in the present study in order to be close to the original procedure. To make self-discrepancies salient, participants were asked to report 10 traits that they actually possessed and 10 traits that they wished they possessed. Participants were then asked to indicate the extent to which each of these traits (actual and ideal) they actually possessed (on a 7-point scale from 1 *not at all* to 7 *totally*) and the extent to which each of these traits they would ideally like to possess (on a similar 7-point scale). While half of the participants were asked to do this before the lexical decision task, the other half completed this task after the lexical decision task. Therefore, only half of the participants had their self-discrepancies explicitly made salient during the lexical decision task assessing suicide thought accessibility.

2.1.3. Lexical decision task

The lexical decision task is a concept accessibility measure widely used in cognitive psychology. The rationale behind this task is that the more cognitively accessible a concept is, the faster the person is to recognize a related word. The task was programmed using Psychopy software (Peirce, 2008). In each trial, after a fixation cross (500 ms), a letter string was displayed on the screen until a key response was pressed by the participant, instructed to indicate as fast as possible whether the letter string was a word (e.g., *ball*) or not (e.g., *blal*) by pressing one of the two allowed keys on the keyboard. Non-words were simple transformations of the words from the task, obtained by switching the position of two adjacent letters (e.g., *chair* and *cahir*). After completing a training session including 10 neutral words and the corresponding 10 non-words, participants were shown 15 neutral words (different from those used during the training session), 5 negative words and 5 suicide-related words, and an equal number of non-words (i.e., 25 non words) in a computer-generated random order different for each participant. Except for the training session, words used in this study were the same as those used by Selimbegović and Chatard (2013). We assessed latencies to correctly recognize suicide-related words, negative words and neutral words (**Table 1**).

2.2. Procedure

The following procedure was approved by the local Internal Ethical Committee of the university where the study was conducted and the participants provided their written consent after reading an information notice about the procedure. All participants were greeted in an experimental room and, after very short instructions from the experimenter, left alone in the room. Participants were randomly assigned to one of the two self-awareness conditions (mirror reflecting the participant's face vs. mirror facing the wall). The experimenter told participants that

Table 1: Lists of words used for the Lexical Decision Task.

Neutral	Negative	Suicide related
Souvent (*Often*)	Triste (*Sad*)	Suicide (*Suicide*)
Livre (*Book*)	Chagrin (*Sorrow*)	Corde (*Rope*)
Bonsoir (*Good Evening*)	Souffrance (*Suffering*)	Veine (*Wrist*)
Vent (*Wind*)	Mauvais (*Bad*)	Pendre (*To hang*)
Haie (*Bush*)	Nul (*Worthless*)	Tentative (*Attempt*)
Train (*Train*)		
Poche (*Pocket*)		
Ballon (*Ball*)		
Le (*The, masculine form*)		
Loi (*Law*)		
Pas (*Not*)		
La (*The, feminine form*)		
Chat (*Cat*)		
Tas (*Pile*)		
Chaise (*Chair*)		

the mirror was there for another experiment conducted by a colleague and that he preferred not to touch his colleague's material. Orthogonally, discrepancy salience was manipulated: participants were randomly assigned to one of the two conditions of discrepancy salience (lexical decision first vs. selves questionnaire first).

Known differences from the original studies are listed in the registration form (https://osf.io/v6bhx). These minor differences are unlikely to substantially influence the results.

3. Results
3.1. Confirmatory (pre-registered) analyses
The following statistical analyses have been pre-registered prior to collecting data. Following Bargh and Chartrand (2000) and as in the original study, recognition latencies longer than 2000 ms were replaced by 2000 ms and recognition latencies associated to wrong answers were discarded. As mentioned earlier, all the tests regarding the mirror effect are directional and hence the reported p-values are one-tailed. Similarly, one-tailed 95% confidence intervals are reported in order to be in line with one-sided testing (Cho & Abe, 2013). As a consequence, confidence intervals include the resulting value in the opposite direction and all the other values toward infinity in the hypothesized direction, hence if the lower bound of the resulting confidence interval is superior to 0, the mirror effect would be significant. Tests relative to the effects that were originally null were kept non-directional and thus two-tailed p-values were reported for these tests.

3.1.1. Pre-registered analyses
As in the original study, latencies to suicide-related words were predicted from mirror exposure, discrepancy salience, and the interaction term between these two variables, and latencies to neutral words were used as a covariate. In accordance with the preregistered exclusion criterion, one participant was excluded from this analysis because his or her score was associated to a studentized residual larger than 3. In this study, the mirror effect was not significant, $t(142) = 0.16$, $p = .57$ (one-tailed), $\eta^2_p < .001$, 95% CI [−0.05, +∞]. Participants in the mirror condition did not recognize suicide related words faster than participants in the control condition ($M = 788$ ms, $SD = 193$ ms, and $M = 779$ ms, SD = 161 ms, respectively). As in the original study, there was no effect of discrepancy salience, $t(142) = 0.13$, $p = .9$, $\eta^2_p < .01$, 95% CI [−0.06, 0.05], and no interaction between the mirror and the selves questionnaire, $t(142) = 0.20$, $p = .85$, $\eta^2_p < .01$, 95% CI [−0.07, 0.09].

We found similar findings when response times to negative words (instead of neutral words) were used as a covariate. The mirror effect was not significant, $t(142) = -0.77$, $p = .22$ (one-tailed), $\eta^2_p < .01$, 95% CI [−0.08, +∞]. Participants in the mirror condition were not significantly faster to detect suicide-related words ($M = 780$ ms, $SD = 193$ ms) than participants in the control condition ($M = 786$ ms, $SD = 161$ ms).[1] In the original study, this effect was marginally significant ($p < .10$, two-tailed). As in the original study, the effect of the self-discrepancy salience manipulation was not significant, $t(142) = -1.20$, $p = .23$, $\eta^2_p = .01$, 95% CI [−0.10, 0.02]. No significant interaction between the mirror condition and self-discrepancies saliency emerged $t(142) = 0.78$, $p = .44$, $\eta^2_p < .01$, 95% CI [−0.05, 0.12].

Hence, the mirror effect failed to replicate following the registered analysis plan.

3.2. Exploratory analyses (not pre-registered)
The use of studentized residuals as an outlier detection method has recently been criticized (Leys, Ley, Klein, Bernard, & Licata, 2013). Indeed, studentization of residuals is computed via the division of the residual by an estimate of the residuals' standard deviation. However,

standard deviations are non-robust parameters sensitive to extreme values. Thus, this method fails to produce a satisfying outlier detection method (see also Rousseeuw, 1990), as it is itself sensitive to outliers. Therefore, Leys et al. (2013) have recently recommended the use of more robust methods to detect outliers, such as the Median Absolute Deviation (MAD).

As suggested by the boxplots presented in **Figure 1**, the outlier detection method preregistered in this replication (studentized residuals) failed to suppress all atypical
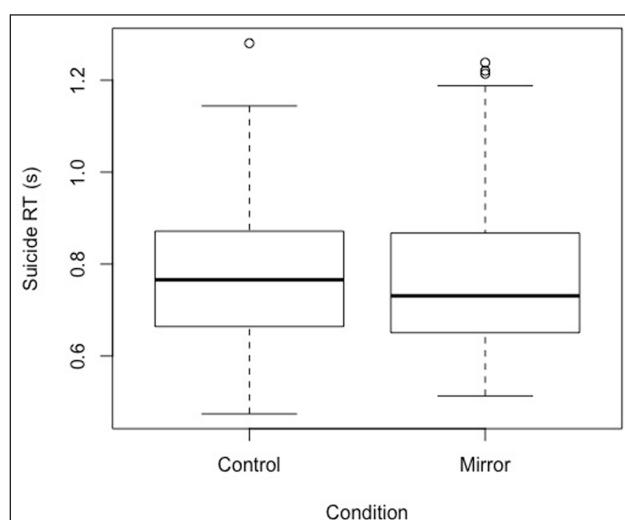

Figure 1: Boxplots of standardized RT to suicide-related words in the control condition (left) ad the experimental condition (right).

observations from the sample. Thus, we decided to conduct complementary analyses using another, more robust, exclusion criterion: the MAD (Leys et al., 2013).

3.2.1. Comparison between the original study and the replication study

In order to more thoroughly examine the original and replicated effects, a comparison was made between the original study and the replication study. We conducted analyses on these two sets of data to investigate how outlier exclusion threshold affects the results. To do this, we observed how the effect sizes in the original study and in the replication varied as a function of the cut-off used to exclude outliers in a multiverse approach (e.g., Steegen, Tuerlinckx, Gelman, & Vanpaemel, 2016). More specifically, we varied the MAD cut-off from 1.5 MAD to 3.5 MAD and observed how partial eta squared evolved when studying the original data with neutral word or negative word latencies as a covariate, and when studying the replication data with neutral word or negative word latencies as a covariate.

When considering the original data, the multiverse analysis showed a mirror effect with neutral word latencies as a covariate. Despite variations of the MAD cut-off value used, the associated p-value varied between $p = .02$ and $p = .10$, with a partial eta squared varying between 2.5% and 5%. With negative word latencies as a covariate, the effect size decreased as the MAD cut-off value decreased (**Figure 2**).

Concerning the replication data, the observed eta squared distributions were different. Whatever the covariate
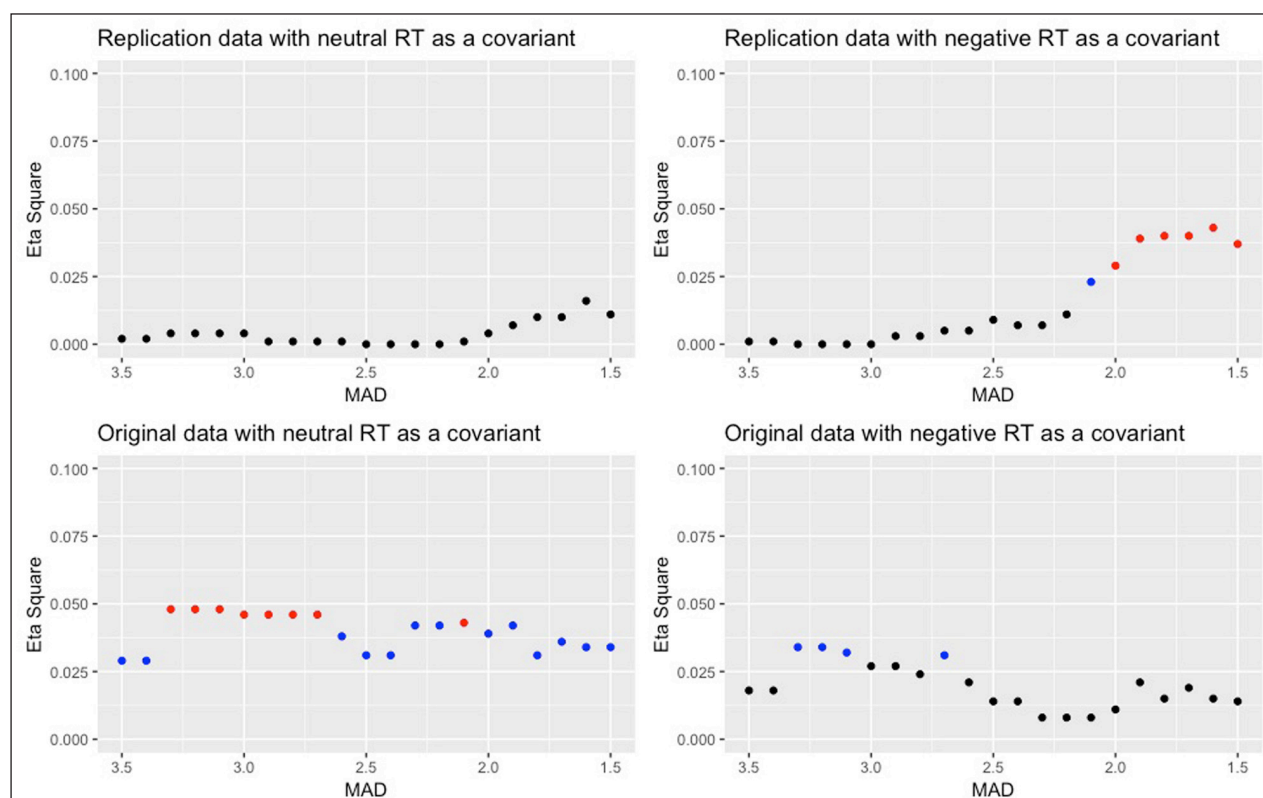

Figure 2: Effect size (η^2_p) as a function of the MAD criterion used to detect outliers. Red dots indicate significance ($p < .05$). Blue dots indicate marginal significance ($p < .10$).

(neutral or negative word latencies), effect size increased as the MAD cut-off became more severe. When negative words latencies were used as a covariate in the replication dataset, and data were excluded using 2MAD as a criterion for outliers' detection, the mirror effect was significant, $t(127) = -1.77$, $p = .04$, $\eta^2_p = .02$, 95% CI [–0.10, +∞]. Participants in the experimental condition were faster at detecting suicide-related words ($M = 728$ ms, $SD = 134$ ms) than participants in the control condition ($M = 767$ ms, $SD = 140$ ms). The mirror effect was therefore replicated in this case and in all the less conservative cases in which the cut-off value for excluding outliers was inferior to 2 MAD.[2]

3.2.2. Equivalence analyses

In order to evaluate if the mirror effect observed in the confirmatory analysis was statistically equivalent to 0, we conducted a TOST (two one sided t-tests). The procedure of the test consists in specifying a smallest effect size of interest (SESOI) and comparing the observed effect to the positive SESOI and the negative SESOI. Two one-sided tests are then conducted to assess if the observed effect size is statistically smaller than the positive smallest effect size of interest and statistically greater than the negative smallest effect size of interest. If the observed effect size is statistically different from these two marks, then we can conclude that the effect is too small to be considered as an effect of interest. However, if one of these two one-sided tests is not significantly different from one of these two marks, then there is no conclusive evidence for equivalence to 0. Conventionally, the reported test is the one of the two one-sided tests that has the largest p-value.

The choice of a SOSEI is a subjective one and depends on a cost/benefit analysis (Lakens, 2017). The investigated effect being mostly of theoretical interest, it is difficult to evaluate its costs and benefits. Recent high-powered meta-analyses reveal that most effect sizes in the field of social psychology are considered to be small (but see Funder & Ozer, 2019). Judging from recent meta-analyses such as Many Labs 4 (Klein et al., 2019), a Cohen's d greater than 0.1 can be accepted as non-trivial (see the pre-registered project of Klein et al., 2019, on the Open Science Foundation website). The original effect size of the mirror effect was a small to medium effect with a lower bound of the confidence interval reflecting a small effect (Cohen's d = 0.43, 95% CI [0.038; 0.827]). A Cohen's d of 0.2 (considered to reflect a small effect, Cohen, 1988; but see Funder & Ozer, 2019 regarding the consequentiality of effect sizes) was chosen as the SESOI. Hence, we specified the lower equivalence bound as a Cohen's d of –0.2 and the upper equivalence bound as Cohen's d of 0.2.

In order to compute the TOST equivalence test, we regressed suicide words RT on neutral and negative words separately RT and saved the residuals from each regression model. This operation allowed us to do the TOST on the variance of suicide words RT that are not explained by neutral words RT on the one hand, and negative words RT on the other hand, consistent with the reported ANCOVA results.

The equivalence test showed that the mirror effect observed when using neutral words RT as a covariant was not statistically equivalent to 0, $t(145.74) = 1.12$, $p = 0.13$. Regarding the equivalence test conducted on the residuals from the regression of suicide words RT on neutral RT, the same conclusions were reached: though participants in the mirror condition did not significantly recognize suicide words faster than participants in the control condition as suggested in the previous ANCOVA, their scores were not significantly equivalent, $t(145.56) = -0.56$, $p = .29$. The results of these equivalence analyses suggest that the replication is inconclusive regarding the evidence for the mirror effect, which remains undetermined in light of the present data.

4. Discussion

In the present study, we attempted to replicate the mirror effect. We expected recognition latencies to suicide-related words to be shorter in the mirror exposure condition than in the control condition, when controlling for neutral words latencies or negative words latencies. These predictions remained unsupported when using the pre-registered outlier detection method in the confirmatory analyses. However, a test assessing the equivalence of the observed effect to a null effect failed to significantly indicate that the mirror effect was equivalent to a null effect (considering d = 0.2 as the smallest effect size of interest). Moreover, an exploratory multiverse analyses showed increasing effect sizes as a function of the decreasing threshold of outlier exclusion, as detected by a robust outlier detection method (i.e, the median absolute deviation, Leys et al., 2013) such that the mirror effect was significant after excluding observations diverging from 2 or less median absolute deviations from the median, but only when using negative words' RT as a covariate. This partial replication raises several interesting questions about the status of the mirror effect, the effect of outliers in a sample, and, more generally, about what allows for concluding that a replication is successful.

4.1. Mixed results concerning the mirror effect

Several large-scale replication projects show that about half of published findings fail to replicate in direct and high-powered replications in psychology (Klein et al., 2018; Open Science Collaboration, 2015; Simons, Holcombe, & Spellman, 2014). These recent studies point out that it is often difficult to replicate published effects. Between the noise inherent to behavioral sciences and the small-sized effects that we often encounter in psychology, observing statistically significant differences is not guaranteed in replication attempts, even when the effect exists in the population. Indeed, one must take into account the inevitable heterogeneity that exists between a study and its replications (Kenny & Judd, 2019), among other factors.

The present replication findings suggest that the original finding might be a false positive. At the same time, equivalence testing does not warrant a conclusion that the effect is equivalent to 0. Also, multiverse analyses show that the effect was significant in some cases, when using a robust method and a severe criterion for detecting outliers. We believe that if the effect exists, the effect size is likely to be smaller than initially thought. In sum, the

study did not provide evidence for a robust mirror effect, but neither did it provide evidence for a null effect (i.e., an effect too trivial to be studied, as defined by a Cohen's d smaller than 0.2). Therefore, further studies using larger samples are needed to establish more reliable estimates of the effect size and a better understanding of the mechanisms involved in this effect, if it exists.

4.2. Detecting outliers in a sample
Outliers are atypical data points that are abnormally different from the "bulk" of observations in a study, and therefore non-representative of the population (Leys, Delacre, Mora, Lakens, & Ley, 2019). There are many ways to define an outlier in a specific data set, as there are many statistical criteria that have been put forward in the literature. Studentized residuals and z-scores are among the most popular ways to detect outliers (Cousineau & Chartier, 2010). However, as underlined by Rousseeuw (1990), these criteria can underperform. The reason for this is that they are based on the sample standard deviation, which is itself a parameter highly sensitive to outliers (Wilcox, 2010). Robust estimators are hence needed to detect outliers. Contrary to studentized and standardized residuals, the median is highly insensitive to outliers (Leys et al., 2013). As one robust estimator, the median absolute deviation (MAD) is particularly relevant in this case, since the classic methods would have failed to detect influential data points (Leys et al., 2013; see also Wilcox, 2017).

How we manage the presence of outliers in a sample is a fundamental aspect of data analysis. However, to date, there is no consensus about which method is the most appropriate and what threshold should be used for detecting and excluding outliers (Leys et al., 2013). In an attempt to optimize the quality of the replication, the hypothesis, method, and statistical analysis were pre-registered. However, what we failed to predict was that excluding outliers on the basis of studentized residuals would not be sufficient to discard all influential data points. Hence, pre-registering a single outlier detection technique might be insufficient. In this view, Leys et al. (2019) recently provided specific recommendations concerning pre-registering and detecting outliers, one of which is to expand a priori reasoning in the registration, in order to manage unpredicted outliers. In our view, this amounts to the option of registering multiple ways to handle outliers. For instance, one could register a *decision tree* regarding the possible ways to handle outliers, as a function of the distribution. For instance, Nosek, Ebersole, DeHaven, and Mellor (2017) mention the possibility to define a sequence of tests and to determine the use of parametric or non-parametric approach according to the outcome of normality assumption tests. In a similar vein, standard operating procedures (SOPs) are procedures more general than decision trees that are shared in a given field of research in order to ground standardization of data handling (e.g., Lin & Green, 2016). The development of such standard procedures applied to outlier detection and exclusion could provide a useful tool for pre-registration.

Developing common, consensual procedures can thus be a solution for dealing with the unpredictable aspects of data, such as the presence of outliers. This would be a controlled, transparent, and probably the optimal manner of handling unpredictability, while suppressing the researchers' degrees of freedom in post-hoc decisions concerning the method used to detect outliers (see Wicherts et al., 2016). In statistics and methodology, as in many fields, a perfect plan does not exist, so it is difficult to offer a perfect solution that fits all studies. In our view, there is a need to define a more general plan of how to handle data, a plan that could fit a large amount of studies. Among the issues that would need to be addressed in such a plan are, for instance, the question of outlier detection/exclusion criterion definition (intraindividually or interindividually), the question of the specific (robust) criterion to be used, and the question of the desired distribution.

5. Conclusion
To sum up, the present replication of the mirror effect yielded mixed findings, since the results depended on the outlier detection method, thereby pointing to a fragile effect. The present findings did not provide much evidence either in favor or against the existence of the mirror effect. They suggest that the mirror effect, assuming that it exists, may be more difficult to detect than previously thought. This underlines the difficulty of conducting well-powered replications and the value of trying to replicate social psychology findings.

Data Accessibility Statement
Materials, participant data, and analysis scripts (R scripts) can be found on this paper's project page on the OSF (https://osf.io/ek2gp/).

Notes
[1] Means are adjusted for the influence of the covariate, hence the difference between the two neutral RT means associated to the two previous ANCOVA.
[2] Though Leys et al. (2013) recommend a 2.5 MAD threshold, they also argue that the use of 2 MAD thresholds can be justified depending on the extent to which outliers are present in the sample.

Competing Interests
The authors have no competing interests to declare.

Author Contributions
· Contributed to conception and design: LS, AC, JM
· Contributed to acquisition of data: JM
· Contributed to analysis and interpretation of data: LS, AC, JM
· Drafted and/or revised the article: LS, AC, JM
· Approved the submitted version for publication: LS, AC, JM

References
Bargh, J. A., & **Chartrand, T. L.** (2000). The mind in the middle: A practical guide to priming and automaticity research. In H. T. Reis & C. M. Judd (Eds.), *Handbook of research methods in social and personality psychology*

(pp. 253–285). New York, NY, US: Cambridge University Press.

Baumeister, R. F. (1990). Suicide as escape from self. *Psychological Review*, *97*(1), 90–113. DOI: https://doi.org/10.1037/0033-295X.97.1.90

Bender, J., O'Connor, A. M., & Evans, A. D. (2018). Mirror, mirror on the wall: Increasing young children's honesty through inducing self-awareness. *Journal of Experimental Child Psychology*, *167*, 414–422. DOI: https://doi.org/10.1016/j.jecp.2017.12.001

Carver, C. S., & Scheier, M. F. (1981). *Attention and self-regulation: A control theory approach to human behavior*. New York: Springer-Verlag. DOI: https://doi.org/10.1007/978-1-4612-5887-2

Chatard, A., & Selimbegović, L. (2011). When self-destructive thoughts flash through the mind: Failure to meet standards affects the accessibility of suicide-related thoughts. *Journal of Personality and Social Psychology*, *100*(4), 587–605. DOI: https://doi.org/10.1037/a0022461

Chatard, A., Selimbegović, L., Pyszczynski, T., & Jaafari, N. (2017). Dysphoria, failure, and suicide: Level of depressive symptoms moderates effects of failure on implicit thoughts of suicide and death. *Journal of Social and Clinical Psychology*, *26*, 1–21. DOI: https://doi.org/10.1521/jscp.2017.36.1.1

Cho, H. C., & Abe, S. (2013). Is two-tailed testing for directional research hypotheses tests legitimate? *Journal of Business Research*, *66*(9), 1261–1266. DOI: https://doi.org/10.1016/j.jbusres.2012.02.023

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences*. New York: Routledge. DOI: https://doi.org/10.4324/9780203771587

Cousineau, D., & Chartier, S. (2010). Outliers detection and treatment: A review. *International Journal of Psychological Research*, *3*(1), 58–67. DOI: https://doi.org/10.21500/20112084.844

Dijksterhuis, A., & Knippenberg, A. V. (2000). Behavioral indecision: Effects of self-focus on automatic behavior. *Social Cognition*, *18*(1), 55–74. DOI: https://doi.org/10.1521/soco.2000.18.1.55

Duval, S., & Wicklund, R. A. (1972). *A theory of objective self awareness*. Oxford, England: Academic Press.

Eitam, B., & Higgins, E. T. (2010). Motivation in mental accessibility: Relevance of a Representation (ROAR) as a new framework. *Social and Personality Psychology Compass*, *4*(10), 951–967. DOI: https://doi.org/10.1111/j.1751-9004.2010.00309.x

Erdfelder, E., Faul, F., & Buchner, A. (1996). GPOWER: A general power analysis program. *Behavior Research Methods, Instruments & Computers*, *28*(1), 1–11. DOI: https://doi.org/10.3758/BF03203630

Fejfar, M. C., & Hoyle, R. H. (2000). Effect of Private Self-Awareness on Negative Affect and Self-Referent Attribution: A Quantitative Review. *Personality and Social Psychology Review*, *4*(2), 132–142. DOI: https://doi.org/10.1207/S15327957PSPR0402_02

Funder, D. C., & Ozer, D. J. (2019). Evaluating Effect Size in Psychological Research: Sense and Nonsense. *Advances in Methods and Practices in Psychological Science*, *2*(2), 156–168. DOI: https://doi.org/10.1177/2515245919847202

Gendolla, G. H. E., & Richter, M. (2010). Effort mobilization when the self is involved: Some lessons from the cardiovascular system. *Review of General Psychology*, *14*(3), 212–226. DOI: https://doi.org/10.1037/a0019742

Heine, S. J., Takemoto, T., Moskalenko, S., Lasaleta, J., & Henrich, J. (2008). Mirrors in the Head: Cultural Variation in Objective Self-Awareness. *Personality and Social Psychology Bulletin*, *34*(7), 879–887. DOI: https://doi.org/10.1177/0146167208316921

Hutton, D. G., & Baumeister, R. F. (1992). Self-awareness and attitude change: Seeing oneself on the central route to persuasion. *Personality and Social Psychology Bulletin*, *18*(1), 68–75. DOI: https://doi.org/10.1177/0146167292181010

Ickes, W. J., Wicklund, R. A., & Ferris, C. B. (1973). Objective self awareness and self esteem. *Journal of Experimental Social Psychology*, *9*(3), 202–219. DOI: https://doi.org/10.1016/0022-1031(73)90010-3

Kenny, D. A., & Judd, C. M. (2019). The unappreciated heterogeneity of effect sizes: Implications for power, precision, planning of research, and replication. *Psychological Methods*. Advance online publication. DOI: https://doi.org/10.1037/met0000209

Klein, R. A., Cook, C. L., Ebersole, C. R., Vitiello, C. A., Nosek, B. A., Chartier, C. R., ... Ratliff, K. A. (2019, December 11). Many Labs 4: Failure to Replicate Mortality Salience Effect With and Without Original Author Involvement. DOI: https://doi.org/10.31234/osf.io/vef2c

Klein, R. A., Vianello, M., Hasselman, F., Adams, B. G., Adams, R. B., Alper, S., ... Nosek, B. A. (2018). Many Labs 2: Investigating Variation in Replicability Across Samples and Settings. *Advances in Methods and Practices in Psychological Science*, *1*(4), 443–490. DOI: https://doi.org/10.1177/2515245918810225

Kruglanski, A. W., Shah, J. Y., Fishbach, A., Friedman, R., Chun, W. Y., & Sleeth-Keppler, D. (2002). A theory of goal systems. In M. P. Zanna (Ed.), *Advances in experimental social psychology*, *34*, 331–378. San Diego, CA: Academic Press. DOI: https://doi.org/10.1016/S0065-2601(02)80008-9

Lakens, D. (2017). Equivalence tests: A practical primer for t-tests, correlations, and meta-analyses. *Social Psychological and Personality Science*, *8*(4), 355–362. DOI: https://doi.org/10.1177/1948550617697177

Leys, C., Delacre, M., Mora, Y. L., Lakens, D., & Ley, C. (2019). How to Classify, Detect, and Manage Univariate and Multivariate Outliers, With Emphasis on Pre-Registration. *International Review of Social Psychology*, *32*(1), 5. DOI: https://doi.org/10.5334/irsp.289

Leys, C., Ley, C., Klein, O., Bernard, P., & Licata, L. (2013). Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median. *Journal of Experimental Social Psychology*, *49*(4), 764–766. DOI: https://doi.org/10.1016/j.jesp.2013.03.013

Lin, W., & Green, D. (2016). Standard Operating Procedures: A Safety Net for Pre-Analysis Plans. *PS: Political Science & Politics*, *49*(3), 495–500. DOI: https://doi.org/10.1017/S1049096516000810

Mor, N., & Winquist, J. (2002). Self-focused attention and negative affect: A meta-analysis. *Psychological Bulletin*, *128*(4), 638–662. DOI: https://doi.org/10.1037/0033-2909.128.4.638

Nosek, B. A., Ebersole, C. R., DeHaven, A. C., & Mellor, D. T. (2017). The Preregistration Revolution. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(11), 2600–2606. DOI: https://doi.org/10.1073/pnas.1708274114

Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, *349*(6251), 1–8. DOI: https://doi.org/10.1126/science.aac4716

Peirce, J. W. (2008). Generating Stimuli for Neuroscience Using PsychoPy. *Frontiers in Neuroinformatics*, *2*, 10. DOI: https://doi.org/10.3389/neuro.11.010.2008

Rousseeuw, P. J. (1990). Robust Estimation and Identifying Outliers. In H. M. Wadsworth (Ed.), *Handbook of Statistical Methods for Engineers and Scientists.* New York, NY: McGraw-Hill, retrieved at https://wis.kuleuven.be/stat/robust/papers/publications-1990/rousseeuw-robustestimation-handbookengineersscient.pdf

Scheier, M. F., & Carver, C. S. (1983). Self-directed attention and the comparison of self with standards. *Journal of Experimental Social Psychology*, *19*, 205–222. DOI: https://doi.org/10.1016/0022-1031(83)90038-0

Selimbegović, L., & Chatard, A. (2013). The mirror effect: Self-awareness alone increases suicide thought accessibility. *Consciousness And Cognition: An International Journal*, *22*(3), 756–764. DOI: https://doi.org/10.1016/j.concog.2013.04.014

Silvia, P. J. (2012). Mirrors, masks, and motivation: Implicit and explicit self-focused attention influence effort-related cardiovascular reactivity. *Biological Psychology*, *90*(3), 192–201. DOI: https://doi.org/10.1016/j.biopsycho.2012.03.017

Silvia, P. J., & Duval, T. S. (2001). Objective self-awareness theory: Recent progress and enduring problems. *Personality and Social Psychology Review*, *5*(3), 230–241. DOI: https://doi.org/10.1207/S15327957PSPR0503_4

Silvia, P. J., & Phillips, A. G. (2013). Self-awareness without awareness? Implicit self-focused attention and behavioral self-regulation. *Self and Identity*, *12*(2), 114–127. DOI: https://doi.org/10.1080/15298868.2011.639550

Simons, D. J., Holcombe, A. O., & Spellman, B. A. (2014). An introduction to registered replication reports at perspectives on psychological science. *Perspectives on Psychological Science*, *9*(5), 552–555. DOI: https://doi.org/10.1177/1745691614543974

Smith, T. W., & Greenberg, J. (1981). Depression and self-focused attention. *Motivation and Emotion*, *5*(4), 323–331. DOI: https://doi.org/10.1007/BF00992551

Steegen, S., Tuerlinckx, F., Gelman, A., & Vanpaemel, W. (2016). Increasing transparency through a multiverse analysis. *Perspectives on Psychological Science*, *11*(5), 702–712. DOI: https://doi.org/10.1177/1745691616658637

Tang, J., Wu, S., & Miao, D. (2013). Experimental test of escape theory: Accessibility to implicit suicidal mind. *Suicide and Life-Threatening Behavior*, *43*(4), 347–355. DOI: https://doi.org/10.1111/sltb.12021

Wicherts, J. M., Veldkamp, C. L. S., Augusteijn, H. E. M., Bakker, M., van Aert, R. C. M., & van Assen, M. A. L. M. (2016). Degrees of freedom in planning, running, analyzing, and reporting psychological studies: A checklist to avoid p-hacking. *Frontiers in Psychology*, *7*. DOI: https://doi.org/10.3389/fpsyg.2016.01832

Wilcox, R. R. (2010). *Fundamentals of modern statistical methods substantially improving power and accuracy.* New York: Springer. DOI: https://doi.org/10.1007/978-1-4419-5525-8

Wilcox, R. R. (2017). *Introduction to robust estimation and hypothesis testing.* Amsterdam: Academic Press. DOI: https://doi.org/10.1016/B978-0-12-804733-0.00010-X

UNIVERSITY of CALIFORNIA PRESS | **Collabra: Psychology**

Collabra: Psychology is a peer-reviewed open access journal published by University of California Press.

OPEN ACCESS