

Failed preregistered replication of mortality salience effects in traditional and novel measures

Bjørn Sætrevik ¹ & Hallgeir Sjøstad ²

1: Operational psychology research group, Department for psychosocial science, Faculty of Psychology, University of Bergen

2: FAIR Insight Team, SNF - Centre for Applied Research at NHH, Norwegian School of Economics

Abstract

Mortality salience (MS) effects, where death reminders lead to ingroup-bias and defensive protection of one's worldview, have been claimed to be a fundamental human motivator. MS phenomena have been supported by more than 400 studies within the "terror management theory" framework, but transparent and high-powered replications are lacking. We performed an Experiment 1 ($N = 101$ Norwegian lab participants) to replicate the traditional MS effect on national patriotism, with additional novel measures of democratic values and pro-sociality. Experiment 2 ($N = 784$ American online participants) aimed to replicate the MS effect on national patriotism in a larger sample, with ingroup identification and pro-sociality as additional outcome measures. Neither experiment replicated the traditional MS effect on national patriotism. The experiments also failed to support conceptual replications and underlying mechanisms on different democratic values, processing speed, psychophysiological responses, ingroup identification, or pro-sociality. This indicates that the effect of death reminders is less robust and generalizable than previously assumed.

Keywords: Mortality salience; death reminders; worldview defence; terror management; replication.

Introduction

Mortality salience.

The concept of mortality salience (MS) refers to a phenomenon where thinking about one's death changes attitudes and behaviour, typically in the form of increased ingroup-bias and behaviour that may serve the role of defending one's cultural worldview. The MS effect has been reported in over 400 experiments since the 1980's, with unusually large effect sizes (for meta-analyses, see: Burke, Kosloff, & Landau, 2013; Burke, Martens, & Faucher, 2010). The dominant theoretical framework to account for MS effects has been the terror management theory (TMT, Greenberg, Pyszczynski, & Solomon, 1986; Pyszczynski, Solomon, & Greenberg, 2015). This theory states that cognitions related to mortality evoke an aversive state of existential anxiety, and motivate to suppress thoughts of vulnerability (proximal defences), or to bolster self-esteem or affirm cultural values to find meaning beyond death (distal defences). It is mainly the MS effects of distal defence, often referred to as "cultural worldview defence", that has been investigated in experiments. In the terminology of TMT and the writings of Ernest Becker (1973), the general idea is that adhering to a cultural worldview can work as a buffer against the fear of death by providing a form of "symbolic immortality". The aim of the current study was to provide a high-powered and pre-

registered replication of the MS effect using traditional outcome measures, and to use novel outcome measures to identify the mechanisms of MS.

Need for replication of mortality salience effects.

Despite the ostensibly solid empirical basis for MS effects from terror management research (Burke et al., 2010), the research tradition has also been called into doubt due to claimed theoretical weakness and non-falsifiability (Martin & Van den Bos, 2014), researcher effects (Yen & Cheng, 2013), failed replications (Trafimow & Hughes, 2012), and contradictory empirical findings (Hart, 2014). Given the recent method reform in psychological research and other fields (Munafò et al., 2017), it should also be noted that there have been few, if any, pre-registered direct replications with open datasets showing MS effects. A review (Burke et al., 2010) is sometimes cited to argue that the MS effect is thoroughly empirically established. However, the review also reveals a great deal of variation in the experiment designs, in terms of different manipulations and outcome measures, whether or not there are "delay tasks" (and their duration and number), and whether the MS effect relies on various covariates. Few of the studies report performing manipulation checks or provide open data. Despite the vast number of studies, no standard experiment approach for producing the MS effect appears to have emerged.

It is noteworthy that there are still few pre-registered replications of the basic MS effect. Hayes and Schimel (2018) performed a series of three experiments, where Study 2 was a pre-registered online experiment. This experiment showed a decrease in self-esteem after performing a word-completion task with death related words. However, the effect only emerged when applying an unregistered exclusion of some of the participants. Pepper and colleagues (2017) failed to replicate the MS effects from a previous study (Griskevicius, Tybur, Delton, & Robertson, 2011), and a recent pre-registered replication of a much cited TMT experiment failed to observe any evidence for a MS effect (Rodríguez-Ferreiro, Barberia, González-Guerra, & Vadillo, 2019). This study also provided an analysis and a discussion of the literature cited by Burke and colleagues (2010), and argued that the distribution of reported effect sizes given the sample sizes did not follow the distribution one would expect from complete reports of a true effect. This makes it an open question to what extent the MS effect can be reproduced and replicated under more restrictive and transparent conditions.

Mechanisms of mortality salience.

MS effects are typically described in relatively general terms (i.e. “threshold for awareness”, and “proximal and distal defences”), without going into details about the cognitive or psychophysiological mechanisms for the effects. As other priming effects, MS may be accounted for by spreading activation in semantic networks (Morewedge & Kahneman, 2010). If so, cognitive representations related to death are closely linked to representations of cultural values in an associative network, so that activation of one part of the network lowers the threshold for activating semantically linked parts of the network (see Arndt, Greenberg, & Cook, 2002, for a similar account). If MS works through such a mechanism, a conceptual replication would be to expect the MS manipulation to lead to increased distraction when a Stroop task presents words related to in-group classification. This may be compared to MS studies that have used lexical decision task and similar measures to estimate death thought accessibility (Hayes, Schimel, Arndt, & Faucher, 2010; Hayes, Schimel, Faucher, & Williams, 2008). Further, it has been argued that MS increases “tension”, “discomfort” or “reluctance” associated with being reminded of death, which proximal or distal defences may reduce (Greenberg, Porteus, Simon, Pyszczynski, & Solomon, 1995; Greenberg, Simon, Pyszczynski, Solomon, & Chatel, 1992). If so, one may expect the MS manipulation to lead to increased psychophysiological activation either during stimulation, or as a residual effect while the MS is in effect (see e.g. Arndt, 1999; Arndt, Allen, & Greenberg, 2001; Rosenblatt,

Greenberg, Solomon, Pyszczynski, & Lyon, 1989). It has also been argued that MS effects may vary with individual differences in cognitive style (Juhl & Routledge, 2010) or political orientation (Burke et al., 2013), which would indicate that one should control for such factors or examine possible interaction effects.

Generalizability of mortality salience.

Although it has been claimed that MS is a fundamental motivator for vast aspects of the human condition (Greenberg et al., 1986; Greenberg, Solomon, & Pyszczynski, 1997), a majority of the studies (72.6% in a meta-analysis, Burke et al., 2010) used similar outcome measures, namely an effect on attitude measures. Moreover, a considerable part of the MS literature have used outcome measures that may be confounded with aggression towards out-group members in the face of threats (such as increase in patriotism or support for the local sports team, see e.g. Stets, 2006; Turner, Oakes, Haslam, & McGarty, 1994). The TMT claim that MS increases cultural values would be better supported if it could be shown that MS increases adherence to cultural values that cannot be construed as out-group aggression.

Finally, although some cross-cultural work has been done (Heine, Harihara, & Niiya, 2002; Routledge et al., 2010), studies in more diverse cultural settings are needed. Cross-cultural studies could show the wider applicability of a MS effect, and could also contribute to exclude competing causal mechanisms. If MS effects could be shown for patriotism and other values outside of North America and Central Europe, this would further support the argument that MS enhances ingroup processes.

Study overview.

As reviewed above, there appears to be some uncertainty around the robustness of MS effects and the underlying mechanism. To address concerns regarding mechanism and generalizability, our Experiment 1 was done in a lab setting and included both the traditional measure of national patriotism and novel measures of democratic values, pro-sociality, Stroop processing and psychophysiology. To address the need for high-powered direct replications, Experiment 2 was conducted as an online experiment of MS effects on a measure of national patriotism, with additional measures of in-group favouritism and pro-sociality. The sample size for main effects in Experiment 1 is about twice as large as the typical effect in the published MS literature, whereas the sample size in Experiment 2 is over 17 times larger than the typical MS study (calculated from the n per analysis cell reported in the meta-analysis of Burke et al., 2010). All hypotheses and analysis approaches for both experiments were pre-registered ahead of data collection, and were performed in accordance with local ethical guidelines.

Experiment 1 Background

Based on the above review, the main aim of Experiment 1 was to replicate the traditional MS effect on national patriotism in a lab setting. As further aims, we also attempted to conceptually replicate the MS effect in novel but theoretically related outcome measures, while controlling for individual cognitive and psychosocial sensitivity to the manipulation. Our pre-registration described four manipulation checks and six hypotheses, listed in Table 2 and discussed below. The pre-registration is available at <https://osf.io/ec4yk>.

To test that the construct of mortality interferes with cognitive processing, we checked whether Stroop RT is higher for death-related words than for neutral words (MC1). To assess whether we succeeded in manipulating MS on a psychophysiological level, we checked whether the MS group increases psychophysiological activation (see MC2 below). Further, we checked that the pro-patriotic (MC3) and pro-democratic essays (MC4) are in fact preferred over the anti essays.

In order to directly replicate the most common type of MS research where MS increases preference for national patriotism, Experiment 1 measured preference for pro and anti-US essays (essays were taken from Rosenblatt et al., 1989, with minor cultural adjustments). Here, we expected (H3) participants in the MS group to show a higher preference for the essay expressing national patriotic values.

Data was collected among a Scandinavian population where democratic values of privacy, citizenship and human rights are mainstream pro-social values. On this background, the patriotic essays were supplemented with essays expressing democratic values when discussing security policy in the aftermath of a terrorist attack. If MS increases the relevance of ingroup cultural values, one would expect (H2) an effect of increased preference for an essay expressing democratic values, compared to an essay expressing anti-democratic values. The novel essays were included to show that MS effects could be shown independently of outgroup aggression.

A possible underlying mechanism of MS may be that existential threats make membership to social groups more important, and thus make people more aware of social categorization. If so, we would expect (H1) that MS activates cognitive structures related to “social categorization”, and thus makes these words more

intrusive on Stroop processing, resulting in longer response times (“RT”).

While most MS studies measure attitudes (as in the two types of essay measures used above), it would be beneficial to supplement this with measures of behavioural intentions. We included a novel measure of pro-sociality that asked participants how they would have shared a hypothetical lottery prize with individuals that were closely or more distantly related to them. This measure was partly inspired Singer’s (2011) concept of a “moral circle” that may include people more or less “distant” from you. In the TMT account, one would expect (H4) the MS group to express a preference for sharing more of the prize with non-relatives, in an attempt to be remembered beyond their physical death.¹ Previous studies have found MS to increase donation to charities (Jonas, Schimel, Greenberg, & Pyszczynski, 2002; Roberts & Maxfield, 2019; Zaleskiewicz, Gasiorowska, & Kesebir, 2015).

Previous studies have claimed that MS effects are more evident in more cognitive flexible individuals (Juhl & Routledge, 2010). We thus expect (H5) the MS effects on the four measures listed above to be enhanced for participants low on “need for closure” (NfC, Federico, Jost, Pierro, & Kruglanski, 2007). Finally, one may expect individual differences in the effectiveness of the MS stimulation, and that this may be indexed by psychophysiological measures. Heart rate variability (HRV, Acharya, Joseph, Kannathal, Lim, & Suri, 2006) may be used as an index of the body’s ability to adapt to the changing demands of the environment. Higher HRV (more high-frequency modulation of heart-beat intervals) has been taken to indicate more physiological adaptability and executive function, while lower HRV has been argued to indicate states of stress and emotional activation (Delaney & Brodie, 2000; Lane et al., 2009). The HRV measurement is unobtrusive after it has been mounted, and is not expected to contribute to fatigue. As the TMT account states that MS leads to an uncomfortable emotional state that is alleviated through either proximal or distal defences. We thus expect (H6) the MS effects on the four measures listed above to be enhanced for participants with lower HRV. See Table 2 for an overview of manipulation checks and hypotheses.

¹ Note that the pre-registrations also included a competing prediction H4b based on the theoretical framework of coalition psychology. Here the MS group participants were expected to express a greater preference to share the prize with individuals with whom they have a genetic relationship, as threats may lead

us to conserve resources to protect relatives. However, as the theoretical reasoning for the competing predictions is weak and neither H4a or H4b were supported, we have excluded this discussion from the current article.

Experiment 1 Methods

Experiment 1 outline

We conducted a laboratory experiment where the predictor variable was MS vs. control manipulation (between subjects, two conditions). The outcome variables were RT for death-related words (representing the extent to which semantic constructs related to mortality are activated by the manipulation), RT for social words (representing the extent to which semantic constructs related to social categorization are activated by the manipulation), preference for pro-democratic essays, preference for pro-patriotic essays, NfC and HRV. The pre-registered experiment procedure and materials can be found online at osf.io/d2zus. Data was collected in September 2015.

Experiment 1 sample

We recruited 101 university students (44 female) through email advertising (pre-registered sample was 100, one participant was replaced during data collection due to non-compliance). All participants were undergraduate psychology students, self-identified as having a Norwegian identity and normal colour vision. Median age group was 22-25 years old. The experiment program randomized participants to the MS group or the control group, without the experimenters knowing who were in which group. Due to an administration error, there were 52 participants in the MS group and 48 participants in the control group. None of the participants were excluded in the data analysis. A sensitivity analysis in the G*power software (Faul, Erdfelder, Lang, & Buchner, 2007), showed that the sample of 100 participants in a between-groups design with two conditions provided 80% power to detect a standardized effect size of $d = 0.5$ (one-tailed, $p < .05$).

Table 1: Experiment procedure stages in Experiment 1.

1	A seven-item scale for «Need for closure», about 1 minute
2	MS manipulation: Two questions where participants were asked to reflect upon and write short string responses about either «death» or «toothache» (depending on condition). 2-3 minutes
3	A «social Stroop» task with half the words related to social categorization, about 5 minutes
4	Pro- and anti-democratic essays (order counterbalanced), each followed by five questions evaluating the essay's content and author, about 5 minutes
5	Pro- and anti-patriotic essay (order counterbalanced), each followed by five questions evaluating the essay's content and author, about 4 minutes
6	Pro-sociality measure, 1-2 minutes
7	A «death Stroop» task with half the words related to death, 2-3 minutes

Experiment 1 procedure

After signing an informed consent form, all participants went through the different stages of the experiment procedure shown in Table 1. The whole experiment took about 25 minutes, and was conducted in sound attenuated testing booths, on desktop computers running the E-Prime experiment presentation software (Psychology Software Tools, 2012), where responses were given with a PC keyboard.

In the beginning of the experiment (stage 1), participants filled out a questionnaire for cognitive style, using a seven item measure of NfC (Federico et al., 2007, translated to Norwegian by the authors).

For the manipulation (stage 2) participants were randomized into two experimental groups, which were manipulated by asking them different questions. The randomization was done by the experiment presentation software and was double blinded for both participants and experimenters. Relying on the traditional manipulation of MS (Rosenblatt et al., 1989), both groups were asked to write down their answers to two short questions presented sequentially, using the traditional instructions of «responding based on gut feelings». The first question asked about thoughts and feelings evoked by «death» for the MS group (or «toothache» for the control group), while the second question asked about what they thought will happen at death and after death (or toothache). A meta-analysis (Burke et al., 2010) has shown that this manipulation is used in 79.8% of the MS studies. A manual inspection showed that all participants provided relevant answers.

A Stroop task was presented (stage 3), with words written in red, blue, green or yellow text against a grey background, and participants were asked to indicate what colour each word was written in as quickly as possible. 50% of the words were related to social categorization (such as «them», «us», «conflict» and «cooperation»), and 50% neutral words matched for letter length and word frequency. There were 200 trials, and the order of presentation for the words was online randomized. Participants responded by using mouse clicks, where the placement of the response boxes switched between trials (this was done to avoid preference for left/right side or centre/lateralized responses, and thus requiring a visual search for the intended response). For each participant, we calculated the ratio of time taken to respond to social words divided by the time on neutral words. Only responses with RT within one SD of the participant's mean were included in the analysis (applying a more lenient criteria of two SD did not significantly change the results).

Next (stage 4 and 5), four brief essays were presented. The first two essays (about 200 words long) were novel for

the study, and presented two opposing views of how Norwegian security policy should be handled in the aftermath of a terrorist attack that happened four years before the data collection. The themes of these essays were whether Norwegian society was essentially safe or under threat, whether extreme viewpoints should be discussed in public or censored, whether the best safeguard against terrorism is social integration and prevention or surveillance and control, and whether terror measures should be balanced against democratic rights or not. The next two essays (about 120 words long) were pro- and anti-patriotic essay that are traditionally used in MS studies (Rosenblatt et al., 1989). In the review of MS experiments (Burke et al., 2010) a wide range outcome variables are used, but the most commonly used is that MS leads to a more favourable evaluation of national patriotic essays (and less favourable evaluation of non-patriotic essays). The essays had been translated into Norwegian by the authors and one aspect changed (from “picking fruits” to “work as store clerk”) to suit the Norwegian setting. All four essays are available online (<https://osf.io/d2zus/>). Whether the pro or anti essays were presented first was counterbalanced between participants. After each of the four essays, participants were asked to rate on a seven-point scale how well they liked the author, how intelligent and knowledgeable they thought the author was, to which extent they agreed with the essay and thought it made an accurate assessment of the issue (higher scores indicating a more positive evaluation). Essay evaluations were calculated as the participant’s average score for the answers to the five questions for each of the four essays (Cronbach’s alpha for pro-democratic essays = .92, for anti-democratic essays = .93, for pro-patriotic essay = .89, for anti-patriotic essay = .91). For each participant, one difference score was calculated between the responses on the pro- and the anti-patriotic essays, and one difference score on the democratic essays (higher scores indicating being more positive to the pro than the anti essays).

Thereafter (stage 6), a novel measure of pro-sociality was applied, where the participants were asked how they would have liked to share a hypothetical lottery prize equivalent to about USD 1.000.000 between (a) their close family (parents and siblings), (b) their relatives (grandparents, uncles and aunts, cousins), (c) their friends and (d) with charities, and (e) what they would keep for themselves and any immediate family. Participants typed in a percentage for how much they wanted to share with each party. All the entered percentages were displayed on screen. The participants were asked to check if they were satisfied with the distribution, and had the option to distribute again. A ratio was calculated of the percentage assigned to the latter two recipients (friends and charity) over the

percentage kept for self, family and relatives $(c+d) / (a+b+e)$.

Finally (stage 7), a second Stroop task (92 trials) was presented, where 50% of the words were related to mortality (e.g. «funeral», «obituary» and «mortal»), with matched neutral words. Presentation, response and calculations were similar as in the first Stroop task (stage 3). This stage was placed last in the experiment, to prevent the presentation of the death-related words to interfere with the assumed MS effect. This is in line with the concern of Hayes and Schimel (2018) that activation of death-related constructs may disrupt the measurement of MS effects.

Throughout the experiment, heart rate was measured from all participants using Polar RS800cx waist sensor and wrist recording unit. We excluded HRV data from participants with signal loss for more than 1/3 of the recording. A 5-minute analysis window was pre-registered to be selected from the middle of the recording (roughly matches when participants evaluate the essays), and the onset was adjusted based on data quality in the window (before identifying participants to their experimental condition). HRV was calculated as the root mean sum of squared differences (RMSSD) of the distance between the peak of each QRS complex (RR beat-to-beat interval, a time-domain analysis). A conventional interpretation (Delaney & Brodie, 2000) is that lower HRV indicates states of stress and emotional activation (but note that this interpretation is sometimes disputed).

Experiment 1 Results

Based on our planned directional predictions, we performed one-tailed null hypothesis testing of our manipulation checks and hypotheses H1, H2, and H3. As the direction of the hypothesis H4 for the pro-sociality task was not clear in the pre-registration, we tested it with a two-tailed test. All analyses were done in the jamovi software (2019). See Table 2 for a summaries of manipulation checks, hypothesis testing and results. Dataset and analyses are available online (<https://osf.io/k28qv/> and <https://osf.io/bgrnu/>).

Manipulation checks.

The first manipulation check (MC1) gave no indication that the semantic construct of «death» was more central for the MS participants. Neither did the second manipulation check (MC2) show a significant effect of MS on psychophysiological activation (HRV). An additional test of HRV at the time of the MS manipulation also failed to find a significant effect of MS (one-tailed $p = .081$). The remaining manipulation checks (MC3 and MC4) showed that the intended essay was preferred in both pairs of essays (pro-Norway and pro-democratic values, respectively),

Table 2: List of manipulation checks and hypotheses, operationalization, tests and extent of support in Experiment 1 (one-tailed *p*-values where not otherwise indicated).

Operationalization	Statistical test	Results
MC1: MS will make semantic constructs related to death more obtrusive on reading		
The MS group will show longer RT for reporting the colour of words with «death related» content compared to words with neutral content.	One-tailed t-test of difference between conditions in ratio of death-word RT to neutral word RT.	Not supported ($N = 99$, MS group $M = 11.9$ ms ($SD = 66$) vs. control group $M = 7.23$ ms ($SD = 48.3$), $p = .345$, $CI = -inf. - 14.7$, $d = -0.08$).
MC2: MS will increase psychophysiological activation		
Participants in the MS group will have lower HRV than the control group.	One-tailed t-test of RMMSD difference between conditions during essay reading.	Not supported ($N = 75$, MS group $M = 44.7$ ($SD = 19.9$) vs. control group $M = 54.1$ ($SD = 31.7$), $p = .065$ in expected direction, $CI = -0.8 - inf.$, $d = 0.35$).
MC3: The pro-democratic essay represents the majority view		
Participants will show a preference for the pro-democratic essay.	One-tailed t-test will show a higher average score on five questions about the pro-democratic essay than average of same questions for the anti-democratic essay.	Supported ($N = 100$, MS group $M = 6.68$ ($SD = 1.29$) vs. control group $M = 3.65$ ($SD = 1.5$), $p < .001$ in expected direction, $CI = 2.73 - inf.$, $d = 1.64$).
MC4: The pro-patriotic essay represents the majority view		
Participants will show a preference for the pro-patriotic essay.	One-tailed t-test will show a higher average score on five questions about the pro-patriotic essay than average of same questions for the anti-patriotic essay	Supported ($N = 100$, MS group $M = 6.45$ ($SD = 1.25$) vs. control group $M = 4.09$ ($SD = 1.5$), $p < .001$ in expected direction, $CI = 2.08 - inf.$, $d = 1.37$).
H1: MS will increase activation of semantic constructs related to social categorization		
Participants in the MS group will have longer RT to the social words than to the neutral words in a Stroop task.	<p>A t-test of experiment group (MS group vs. control group) on Stroop effect for social words as outcome variable.</p> <p>To test the H5a prediction, the same relationship was also tested with a two-way ANOVA with an added interaction of split-half scores of NfC.</p> <p>As predicted by H6a prediction, the same relationship was also tested with a two-way ANOVA with an added interaction of split-half scores of HRV.</p>	<p>Not supported ($N = 99$, MS group $M = -2.29$ ms ($SD = 40.8$) vs. control group $M = 11.6$ ms ($SD = 58.5$), $p = .09$, $CI = inf. - 30.7$, $d = .28$).</p> <p>H5a not supported (two-tailed $p = .21$).</p> <p>H6a not supported (two-tailed $p = .77$).</p>
H2: MS will increase support for democratic values		
Participants in the MS group will show a higher preference for the democratic essay compared to the anti-democratic essay.	<p>A t-test for experiment group (MS group vs. control group) on pro-/anti-democratic essays as outcome variable.</p> <p>To test the H5b prediction, the same relationship was also tested with a two-way ANOVA with an added interaction of split-half scores of NfC.</p> <p>To test the H6b prediction, the same relationship was also tested with a two-way ANOVA with an added interaction of split-half scores of HRV.</p>	<p>Not supported ($N = 100$, MS group $M = 52$ ($SD = 1.89$) vs. control group $M = 48$ ($SD = 1.82$), $p = .29$, $CI = -inf. - 0.41$, $d = 0.11$).</p> <p>H5b not supported (two-tailed $p = .89$).</p> <p>H6b not supported (two-tailed $p = .24$).</p>
H3: MS will increase national patriotism		
Participants in the MS group will show a higher preference for the patriotic essay compared to the anti-patriotic essay.	<p>A t-test for experiment group (MS group vs. control group) on patriotic essays as outcome variable.</p> <p>To test the H5c prediction, the same relationship was also tested with a two-way ANOVA with an added interaction of split-half scores of NfC.</p> <p>To test the H6c prediction, the same relationship was also tested with a two-way ANOVA with an added interaction of split-half scores of HRV.</p>	<p>Not supported ($N = 100$, MS group $M = 2.33$ ($SD = 1.75$) vs. control group $M = 2.4$ ($SD = 1.73$), $p = .58$, $CI = -inf - 0.65$, $d = .04$).</p> <p>H5c not supported (two-tailed $p = .48$).</p> <p>H6c not supported (two-tailed $p = .55$).</p>

H4: MS will affect the degree of pro-sociality		
The amount shared with friends and charities relative to the amount shared with family will be different for the MS group.	A t-test for experiment group (MS group vs. control group) on the ratio of giving to friends and charities as outcome variable.	Not supported ($N = 100$, $M = 0.44$ ($SD = 0.71$) vs. $M = 0.25$ ($SD = 0.31$), two-tailed $p = .09$, $CI = -inf. - -0.01$, $d = 0.35$).
	To test the H5d and H5e predictions, the same relationship was also tested with a two-way ANOVA with an added interaction of split-half scores of NfC.	H5d not supported (two-tailed $p = .31$).
	To test the H6d and H6e predictions, the same relationship was also tested with a two-way ANOVA with an added interaction of split-half scores of HRV.	H6d not supported (two-tailed $p = .33$).

confirming that the essay measure was successful at creating a cultural ingroup versus outgroup scenario.

Confirmatory analyses

The planned tests of H1, H2 and H3 were performed with one-tailed t-tests in the direction stated in the pre-registration. Each were followed up with two-way ANOVAs, where the split-half classification of high/low NfC and high/low HRV was added to test for interactions (respectively corresponding to H5 and H6 hypotheses in the pre-registration). This approach is functionally equivalent to the GLM approaches described in the pre-registration. The planned tests for H1 showed no main effect of MS on the «social Stroop» test, nor any interaction with NfC or HRV. H2 showed no main effect of MS on rating of democratic essays, nor any interaction with NfC or HRV. H3 showed no main effect of MS on the rating of patriotic essays, nor any interaction with NfC or HRV. Thus, Experiment 1 did not find any support for the primary hypotheses of the MS effects found in the TMT literature.

On the measure of pro-sociality, the MS group stated they would share more (23.1%) with friends and charities than the control group (16.9%). An effect in that direction would support the H4 prediction derived from TMT that MS should increase preference for sharing with non-relatives. However, the effect was not significant in the two-tailed test prescribed by the bidirectional hypothesis in the pre-registration. Nevertheless, the difference in means between the groups justifies a further examination of this measure, which is performed in Experiment 2. There were no significant interactions of NfC or HRV on the pro-social task.

Experiment 2 Background

The main aim of Experiment 2 was to directly replicate the traditional effect of MS on national patriotism in an American sample. As additional aims we wanted to test conceptual replication of MS effects on ingroup identification and pro-sociality. The pre-registration for Experiment 2 is available at <https://osf.io/d26fq>.

The first hypothesis (H1) tested whether MS effects involve mechanisms of ingroup identification and group

membership. This corresponds to the mechanism tested in H1 in Experiment 1 (the “social Stroop test”), but the current test has higher face-validity. Moreover, the test is an attempt to replicate an effect of MS leading to higher ingroup identification previously found in an Italian sample (Castano, Yzerbyt, Paladino, & Sacchi, 2002).

The second hypothesis (H2) was intended to directly replicate the traditional lab experiments, where participants who write short answers to questions about death showed increased patriotism later in the study (Greenberg, Pyszczynski, Solomon, Simon, & Breus, 1994; Greenberg et al., 1992; Simon et al., 1997). We attempted to make both manipulation and effect measure as similar as possible in an online experiment to the traditional experiments.

The third hypothesis (H3) was intended to further explore the effects of MS on the novel measure of pro-sociality used in Experiment 1. This measured generosity towards people outside your family, and is thus an indicator of pro-sociality. Previous studies have shown MS to lead to increased pro-sociality (Jonas et al., 2002; Roberts & Maxfield, 2019; Zaleskiewicz et al., 2015).

In order to maintain similarity with the research we want to replicate, we also included a mood measure that functioned as a delay task after the MS manipulation, and before the three outcome measures. Finally, we also measured political orientation in order to perform a pre-registered exploration of whether political orientation moderates the predicted effect of H1, H2 or H3.

Experiment 2 Methods

Experiment 2 outline

We conducted a high-powered online experiment on a total sample of 800 US participants. As in Experiment 1, the predictor variable was MS versus control. The outcome variables were American ingroup identification, national patriotism, and the same measure of pro-sociality as in Experiment 1.

Experiment 2 sample

A total of 803 US participants signed up for this study in exchange for \$0.50 USD. After excluding 19 duplicate and one incomplete responder, the final sample consisted of

784 participants (age $M = 38$, 61% female). Data was collected on July 12th, 2019. A sensitivity analysis in the G*power software showed that a sample of this size provides 90% power to detect a standardized effect size of $d = .23$ or larger in a two-tailed test ($p < .05$). As noted earlier, this sample is more than 17 times larger per analysis cell than the typical MS experiment in the TMT literature.

Experiment 2 procedure

The Amazon Mechanical Turk online platform (Buhrmester, Kwang, & Gosling, 2011; Hauser & Schwarz, 2016) was used to recruit participants for a study about “personality and attitudes”, and the experiment was programmed and administered on the Qualtrics platform (available as online materials: <https://osf.io/jm4uh/>). On average participants spent just over 8 minutes on the experiment. See Table 3 below for an overview of experiment procedure.

When clicking through to the survey, participants were randomized to receive the MS or control manipulation. At the beginning of the experiment (stage 1), the MS and control group were asked to write brief answers to the same two questions as in Experiment 1 about either “death” or “toothache”, respectively. The two questions were presented on separate pages, and answers were written in an empty text box beneath each question. To avoid that participants could simply click their way through the survey without responding, the answer to each of the two questions had to be at least 15 characters long before they could proceed. As manipulation checks, we (MC1) verified that the pro-USA essays were in fact preferred over the anti-USA essays. In addition, we (MC2) manually verified that participants had in fact provided meaningful answers to the manipulation questions about death (pre-registration of this analysis: <https://osf.io/sw6md>). While this approach to a manipulation check has weaknesses, it should be noted that manipulation checks even to this extent are rarely reported in the published MS literature.

On the next screen (stage 2), there was a delay task of answering 20 questions about current mood (PANAS; Watson, Clark, & Tellegen, 1988). Participants were asked to rate the extent they felt “interested”, “distressed”, “excited” etc. on a five-point scale from “Not at all” to “Extremely”. This task was included to maintain similarity to traditional MS experiments, where the account of MS effects emphasizes the need for a delay task between the manipulation and the outcome variables. The same task is used in 47.7% of the MS literature (Burke et al., 2010), most typically as the only delay task.

Thereafter (stage 3) participants were presented with a screen with five statements about their American identity, based on the group identification scale (Doosje, Ellemers, & Spears, 1995). Each item was rated on a seven-point

scale from “Not at all” to “Totally”. The statements were: “*I perceive myself as an American*”, “*I feel strong ties with other Americans*”, “*Being an American does not mean much to me*” (reversed), “*I identify with American people*”, and “*Being an American has nothing to do with my identity*” (reversed). Cronbach’s alpha for the responses was .84.

Next (stage 4), national patriotism was measured with one pro-USA and one anti-USA essay presented separately. These were the traditional essays used in MS experiments (see e.g. Rosenblatt et al., 1989) and similar to the essays used in Experiment 1, except that they had not been translated and adapted to fit a Norwegian context. All participants read both the pro-USA and the anti-USA essays in counterbalanced order. After reading each essay, participants answered the same five questions as in Experiment 1 (Cronbach’s alpha = .91 for pro-USA, .95 for anti-USA). To get a score for national patriotism, the average score on the anti-USA essay was subtracted from the average score on the pro-USA essay (higher scores indicate higher patriotism).

Thereafter (stage 5), the same pro-social task of pro-sociality as in Experiment 1 was applied (value of lottery win set to USD 1.000.000). In a sequential list, participants entered the percentage they would like to share with (a) self or immediate family, (b) close family, (c) extended family, (d) friends, and (e) charities. The summed percentage was shown beneath, and the sum had to be 100% in order to continue the experiment. As in Experiment 1 and as pre-registered, an index for pro-sociality was calculated as $(d+e) / (a+b+c)$, in which a higher number indicates a higher level of pro-sociality.

At the end of the experiment (stage 6), participants reported age and gender, and then a single-item measure of political views as a potential moderator variable. The question “*In general, what would be the most accurate description of your political views?*” was answered on a 7-point scale (marked with 1 = Very left-wing/ liberal, 4 = centrist/ moderate, 7 = Very right-wing/ conservative). The responses used the full range, with a central tendency in the middle ($M = 3.64$, $SD = 1.8$).

Table 3: Experiment procedure stages in Experiment 2.

1	MS manipulation: Two questions where participants were asked to reflect upon and write short string responses about either «death» or «toothache» (depending on condition), 2-3 minutes
2	Delay task: Indicate current mood, 20 questions, about a minute
3	DV1: Ingroup identification, 5 questions, about half a minute
4	DV2: National patriotism, ratings of pro-USA and anti-USA essays (order counterbalanced), about 3 minutes
5	DV3: Pro-social task, 1-2 minutes
6	Moderator: variable: Political orientation, a few seconds

Table 4: List of manipulation checks and hypotheses, operationalization, tests and extent of support in Experiment 2 (two-tailed p-values).

Operationalization	Statistical test	Results
MC1: The pro-USA essay will be preferred over the anti-USA essay (across conditions)		
Participants will show a preference for the pro-patriotic essay.	One-tailed t-test will show a higher average score on five questions about the pro-patriotic essay than average of same questions for the anti-patriotic essay	Supported ($N = 784$, MS group $M = 5.4$ ($SD = 1.04$) vs. control group $M = 4.4$ ($SD = 1.62$), $p < .001$ in expected direction, $d = 0.59$).
MC2: The manipulation instructions were followed		
Participants will provide meaningful answer to the manipulation questions	Manual classification of all 800 responses	98% provided relevant responses
H1: MS will increase ingroup identification		
Participants in the MS group will have to a larger degree identify as Americans.	<p>A t-test of experiment group (MS group vs. control group) on ingroup identification score.</p> <p>Regression of political views on ingroup identification.</p> <p>Moderation of political views on relationship between mortality salience and ingroup identification.</p>	<p>Not supported ($N = 784$, MS group $M = 5.2$ ($SD = 1.38$) vs. control group $M = 5.28$ ($SD = 1.39$), $p = .46$, $d = 0.05$).</p> <p>More conservative participants showed higher ingroup identification for ($t = 9.67$, $p < .001$).</p> <p>No moderation effect of political views ($Z = -0.8$, $p = .425$).</p>
H2: MS will increase national patriotism		
Participants in the MS group will show a higher preference for the patriotic essay compared to the anti-patriotic essay.	<p>A t-test for experiment group (MS group vs. control group) on patriotic essays as outcome variable.</p> <p>Regression of political views on patriotism.</p> <p>Moderation of political views on relationship between mortality salience and patriotism.</p>	<p>Not supported ($N = 784$, MS group $M = 1.15$ ($SD = 1.99$) vs. control group $M = 1.17$ ($SD = 1.95$), $p = .91$, $d = 0.01$).</p> <p>More conservative participants showed higher patriotism ($t = 11$, $p < .001$).</p> <p>No moderation of political views on patriotism ($Z = -0.04$, $p = .97$).</p>
H3: MS will increase pro-sociality		
Participants in the MS group will state that they would share more of a hypothetical money prize with friends and charities relative to the amount shared with family and relatives.	<p>A t-test for experiment group (MS group vs. control group) on the ratio of giving to friends and charities as outcome variable.</p> <p>Regression of political views on pro-sociality.</p> <p>Moderation of political views on relationship between mortality salience and patriotism.</p>	<p>Support for less sharing in MS group ($N = 781$, $M = 0.177$ ($SD = 0.267$) vs. control group $M = 0.234$ ($SD = 0.234$), $p = .036$, $d = .15$).</p> <p>No effect of politics on pro-sociality ($t = 1.62$, $p = .106$).</p> <p>No moderation effect of political views on sharing ($Z = 1.5$, $p = .135$).</p>

Experiment 2 Results

In accordance with the pre-registration we tested H1, H2 and H3 as t-tests against a two-tailed p-value of .05. In addition, we tested the effects of political orientation on the outcomes as simple regressions, and we tested their moderating effect on the relationship between MS and outcomes. All analyses were done in the jamovi software (2019), using the “medmod” module for the moderation analyses. See Table 4 for results from the pre-registered confirmatory hypotheses tests (H1, H2 and H3). Dataset and analyses are available online (at <https://osf.io/mejnt/> and <https://osf.io/zpn92/>).

The results showed no significant difference between the MS and the control group on ratings of ingroup identification (H1). Further, there was no significant difference between the MS and control group on the focal outcome measure of national patriotism rating (H2). Thus, Experiment 2 did not show any support for the primary hypotheses about the effect of death reminders, as derived from TMT and previous research. On the pro-sociality measure there was a small but significant difference in the opposite direction of the H3 prediction ($d = .15$, $p = .036$), of the MS group sharing less (12.7% of the amount) than the control group (14.8%) with friends and charities.

As suggested in the pre-registration, we explored the possible moderator effect of political orientation on the outcome variables. Political orientation was significantly correlated with two of the outcome measures, indicating that more conservative participants identified more as Americans and showed higher national patriotism, while there was no effect on pro-sociality. However, there was no significant interaction between MS and political orientation on neither of the three outcome measures (ingroup identification, national patriotism and pro-sociality), thus showing no difference in how conservative and liberal participants responded to the MS manipulation.

Discussion

The aim of the current study was to directly replicate the effect of MS on attitudes to national patriotic essays, and to conceptually replicate the effect on other measures to explore mechanisms and boundary conditions, using pre-registered lab and online experiments with a combined sample of 884 participants from two different countries. Despite our best efforts, we failed to both directly and conceptually replicate the MS effects. Neither did we find indications of the assumed mechanism of MS (word processing times, psychophysiology or ingroup identification). The second experiment showed a small but significant effect on pro-sociality, where MS led to reduced pro-sociality. This effect is in the opposite direction from the prediction derived from traditional MS theories and previous research. We were thus unable to obtain any empirical support for direct or conceptual replication of the MS effect or its assumed mechanisms on any of the outcome measures. As opposed to most previous research on MS and TMT, we performed manual verification for both experiments that the manipulation was adhered to (that the MS group in fact wrote relevant answers related to “death” themes), and provide this and all other outcomes in public datasets. Based on this analysis, we see it as unlikely that the null-results can be explained by a failure to manipulate the independent variable (death awareness vs control). The results are further discussed below.

No direct replication of MS effect on patriotism

Both experiments failed to directly replicate the traditional MS effect on national patriotism, using the typical essay measure in both a Norwegian (H3 in Experiment 1) and in an American context (H2 in Experiment 2). Although our manipulation checks confirmed that the patriotic essay reflected the dominant cultural values in our samples, the MS treatment did not increase this preference. This result opposes much of the published MS literature, typically

described in terms of the TMT framework (see e.g. Burke et al., 2013).

Since the aim of Experiment 1 was a conceptual rather than direct replication, a Stroop task and a novel essay aimed at testing theoretical extension was performed before the patriotic essay. We assumed that these measures would be sufficiently indirect to not interfere with the effect on patriotic essays. Nevertheless, the high-powered Experiment 2 had the patriotic essays immediately after the delay task, which would constitute a direct replication of the prototypical MS study.

No direct replication of MS effect on ingroup identification

As a straight-forward test of the mechanism assumed cause the MS effect, Experiment 2 (H1) tested whether the manipulation increased social identification with the larger ingroup (i.e., being an American). There was no effect on this measure, thus failing to support what has been claimed to be the fundamental mechanism behind MS effects in the TMT. This also constitutes a failed direct replication of a previous Italian study using similar measures (Castano et al., 2002). To our knowledge, no other studies have directly tested this assumed mechanism of MS on ingroup identification, but have instead tested the effect that MS has on expressing the ingroup’s values.

No conceptual replication of MS effect in novel essays

We constructed essays for measuring MS effects on preferences for democratic values for use in Experiment 1. A manipulation check (MC3) confirmed that the pro-democratic essay expressed values that were dominant in the sample, as it was rated higher on average than the anti-democratic essay. Our theoretical extension (conceptual replication) of the traditional MS effect was that we expected that MS would increase the expression of democracy. However, the H2 test did not show a MS effect of increased preference for the democratic essay.

Experiment 1 thus failed to demonstrate the MS effect to transfer to a novel and culturally adapted measure (support of democratic values). Although this replication in Experiment 1 contain novel aspects, the literature often describes MS effects as being universal across cultures, and that the effect has wide-reaching consequences for most aspects of human life. Further, reviews have shown MS effects in a wide range of outcome variables (Burke et al., 2010), and has been shown to transfer to defence of cultural values both American, European and non-Western societies (e.g. Heine et al., 2002; Routledge et al., 2010).

No effect of MS on word processing speed

There was no effect in the Experiment 1 manipulation check of MS increasing Stroop processing times for death-related words (MC1). This fails to support the claims of the TMT (see e.g. Arndt et al., 2002; Arndt, Greenberg,

Solomon, Pyszczynski, & Simon, 1997), as there was no indication that MS manipulation made concepts related to mortality more accessible for the participant.

Neither did Experiment 1 find a MS effect on Stroop processing times for words associated with social categorization (H1). This opposes the expectation derived from TMT that MS leads to making social identification (or cultural belonging) more relevant to find meaning beyond physical death. We propose three possible explanations for the lack of Stroop effects. Either (1) the MS effect is not a real phenomenon (at least not in the form described in the literature), or (2) MS does not work through a basic cognitive mechanism of spreading activation in a conceptual network, and can thus not be measured with a Stroop task, or (3) our Stroop methodology was not suitable to register changes in availability of cognitive constructs. Using a computer mouse for Stroop responses may produce some random variation in response times, but this should be compensated for by the high number of Stroop trials.

To our knowledge, no former experiments have tested a direct effect of MS on processing speed of death-related concepts, or an indirect effect on concepts such as social words that are assumed to be causally linked with mortality. Some studies (e.g. Gailliot, Schmeichel, & Baumeister, 2006) have suggested that MS slows down Stroop processing in general, but without using Stroop words with relevant/irrelevant content. This leaves us without an established framework to evaluate whether Stroop is a suitable approach to test the mechanisms assumed to underlie the MS effects. We encourage further testing of this approach in future studies.

No MS interaction effect on psychophysiology

The TMT assumes that MS leads to an uncomfortable state that motivates the affirmation of one's cultural values (Arndt et al., 2001; Delaney & Brodie, 2000; Henry, Bartholow, & Arndt, 2010; Lane et al., 2009; Routledge et al., 2010; Schuler, Mlynski, & Wright, 2017; Silveira et al., 2013). We assumed that such a change of state could be detectable in psychophysiology. However, one of the manipulation checks in Experiment 1 (MC2) failed to show a significant effect of decreased HRV in the MS condition. An effect of MS on HRV may have been indicated (one-tailed $p = .065$), but did not meet our pre-registered alpha level. A psychophysiological activation effect may have been obscured by individual variation, measurement noise or analysis choices.

It is possible that different HRV analysis approaches (such as different artefact smoothing or different analysis windows) could have indicated effects. However, note that there were no effects in an additional testing window (during MS stimulation), nor in the interactions between

MS and HRV on any of the outcome measures. Given the high volatility of HRV analysis to analysis flexibility, we chose to not explore the data outside of our planned analysis. Note that research on physiological indicators of MS appear to have shown mixed results, also within the TMT literature (Arndt, 1999; Rosenblatt et al., 1989).

Effect of MS to decrease pro-sociality

Both experiments included our novel money-sharing measure of whether MS increased pro-sociality. A conceptual extension of TMT was that MS should lead to increased generosity to people outside of one's family, in order to be remembered beyond one's death (building on Jonas et al., 2002; Roberts & Maxfield, 2019; Zaleskiewicz et al., 2015). Experiment 1 showed a non-significant tendency of MS leading to increased money-sharing with friends and charity. However, the more robust test in Experiment 2 showed a significant effect in the opposite direction, of less sharing in the MS group. This is difficult to align with the TMT account, as being less generous with friends and strangers cannot be seen to be a culturally dominant value. One could perhaps come up with ad hoc explanations for why this measure showed opposite effects in the two studies (e.g. different situational priming, different sample populations). However, in order to adhere to the hypo-deductive approach of null-hypothesis significance testing, we would advise extreme caution in interpreting a non-predicted result as supportive evidence for a given hypothesis. Further, one should keep in mind that the pro-social task is novel and unestablished.

Evaluation of replication effort

In an attempt to account for the current null-findings, one may point to differences between the current experimental design and previous studies. For example, there was no explicit delay task before the target measure in Experiment 1, which previous research suggest may be necessary for MS effects to occur (Greenberg, Arndt, Simon, Pyszczynski, & Solomon, 2000). On the other hand, one could argue that the Stroop test and the democratic essays fill this role, as they take place between the MS manipulation and the patriotic essays measure and are ostensibly unrelated to both mortality and patriotism. Further, Experiment 2 used the most common type of delay task (a 20-item mood measure) and still failed to show a MS effect on national patriotism or ingroup identification in a large American sample. The opposite argument could also be made, that the number of tasks led to a fatigue that could hide a MS effect. However, in both experiments the patriotic essays were among the first measures following the manipulation and delay task, to ensure a sensitive test of the primary hypothesis. As mentioned above, a review (Burke et al., 2010) argued that MS effects have been showed across a number of experiment procedures using

zero, one, two or three delay tasks. As the TMT claims that MS is a fundamental motivator for human behaviour, it would be surprising if it relies on exact repetition of minor variations in experimental procedure discussed here. We would argue that Experiment 2 adheres to the crucial design features in the published literature, in having one delay task immediately preceding the central outcome measures.

Following previous research, our experiments also tested whether the outcome variables showed interactions with cognitive style, political orientation, or the manipulation's effectiveness (although these tests are lower powered than the main effect tests). These tests failed to provide any further support for the MS effects.

In terms for statistical power, Experiment 1 had 50 participants in each condition, while Experiment 2 had 400 in each condition. According to a meta-analysis of the MS effect in more than 400 experiments (Burke et al., 2010), the average effect size was $d = 0.82$. When dividing the stated N for the studies by the first and second predictor variable, resulted in a median sample size per condition of $n = 23.3$. Another meta-analysis of the effect of MS on political attitudes (Burke et al., 2013) across 49 experiments stated an average effect size of $d = 1.15$, and a similarly calculated median sample size of $n = 25.9$. Some of the studies may also have had additional, undeclared predictors (John, Loewenstein, & Prelec, 2012). Thus, the sample size in Experiment 1 is at least twice as large as the typical MS study, while Experiment 2 is over 17 times larger than the typical MS experiment. The sample size of Experiment 1 provided 80% power to detect a MS effect of $d = 0.5$ using a one-tailed test, and was therefore more than sufficient to detect effects of the magnitudes described by Burke and colleagues (2010). The sample size of Experiment 2 is far larger than recent recommendation of using at least 2.5 times the original sample to ensure high-powered replications (Simonsohn, 2015), and a sensitivity analysis showed that it provided 90% power to detect a MS effect of $d = .23$ or larger, using a two-tailed test. Thus, even if the true effect should be only one-third of the smallest meta-analytic effect of MS ($d = .75$), our Experiment 2 would still have a 90% chance to detect the main effects.

Although participants with right-leaning political views were more supportive of the national-patriotic essay than left-leaning liberal participants, exposure to the death reminder (MS) did not lead to increased national patriotism in this sub-group either (i.e. the same null-effect for this sub-sample as for the overall sample).

In order to achieve this level of statistical power, Experiment 2 recruited a large online sample using Amazon's Mechanical Turk, while most of the previous literature has been done in-person in physical labs.

Although there are some concerns about the representativeness of convenience samples at Mechanical Turk and similar sites, closer examination has indicated that these samples may be more representative for the general population than the student samples typically used in lab experiments (Buhrmester et al., 2011). Moreover, some studies have indicated that online participants at Mechanical Turk pay closer attention to study instructions than student samples (Hauser & Schwarz, 2015) and provide comparable data quality (Kees, Berry, Burton, & Sheehan, 2017). In our view, this suggests that in relatively short and focused survey experiments online samples can be a valuable trade-off to student samples in order to achieve high-powered studies. There is little in the established account of MS phenomena to indicate that it should be restricted to in-person studies. However, if it should turn out that transparent and high-powered in-person lab procedures can reliably produce the MS effect while online studies fail to do so, that boundary condition should be implemented in the theoretical account of the phenomenon.

A possible explanation for the current null-finding is that the original effect is less robust than previously assumed. As with the majority of psychological research preceding the recent awareness of fundamental methodological issues, most MS research has been conducted in a non-transparent way without pre-registration or open data (Munafò et al., 2017; Simmons, Nelson, & Simonsohn, 2011). This makes it difficult to assess the extent of unpublished results and undisclosed flexibility in design and analysis in the MS literature. The current pre-registered null-findings in a controlled lab study and in a large-scale online sample indicate that the MS effect may not be a robust and replicable phenomenon, despite the high number of past publications (Burke et al., 2013; Burke et al., 2010). At the very least, the current null-findings emphasizes the need for high-powered, pre-registered and transparent replication of the traditional MS effect.

Conclusion

The current study with a Norwegian lab experiment ($N = 101$) and an American online experiment ($N = 784$) aimed to directly replicate an often reported MS effect on national patriotism, and a previous reported MS effect on ingroup identification. The study further aimed to conceptually replicate the MS effect on support for democratic values, and to explore a potential MS effect on a novel measure of pro-sociality. All these efforts failed to support the predicted MS effects. One of the experiments found a significant MS effect of decreased pro-sociality, but this effect is in the opposite direction of the hypothesis derived from the established literature. The lab experiment was also

unable to find any effect of MS on processing speed of concepts related to death or social categorization. An increased psycho-physiological activation from MS may have been indicated, but failed to reach the cut-off for one-tailed significance in two different analysis windows. We also failed to support interaction effects derived from reasonable interpretations of the MS literature.

Some methodological shortcomings are discussed above. One could claim that while being more transparent and better powered than most of the cited literature, Experiment 1 is nevertheless underpowered and has a rather complex experimental design. However, Experiment 2 can be interpreted alone as a high-powered attempt at directly replicating the central MS effect. Given the claim that MS effects are robust and should generalize across a variety of settings and outcome measures (Burke et al., 2013; Burke et al., 2010; Pyszczynski et al., 2015), it is noteworthy that both our attempts at pre-registered replications of the traditional MS effect failed. If one would like to argue that past literature offers empirical support for the MS effect, one should define the necessary and sufficient conditions to produce the effect (e.g. type and duration of delay task, lab or survey-based data collection, which covariates are necessary), and one should only count studies that fulfil these conditions as supportive of the effect.

In our view, the current results show that the basic MS effect is more difficult to reproduce than previously assumed. It is possible that a MS effect hinges on methodological quirks, specific samples or boundary conditions that have not been reported or identified in previous research. We welcome further research on this, but we will view the proposed MS effect with scepticism until such conditions are identified. We actively encourage attempts to replicate the null-findings from the current experiments. Divergent findings could help to identify boundary conditions for the effect (Nosek & Errington, 2019).

Note that independently of our study, a recent “Many Labs 4” project (Klein et al., 2019, December 11) has attempted to replicate the MS effects across 21 different labs ($N = 2.220$). These pre-registered experiments failed to replicate the original studies, both with and without original author involvement. These findings mirror the current results, and additionally indicate that methodological expertise in study design is not a likely explanation for the null-findings, nor that the MS effect can be reliably reproduced in lab studies.

We find it uncontroversial that avoidance of death can be a powerful motivator, and that human psychology is embedded with instincts to favour and support the ingroup. However, it is less obvious that an abstract

awareness of mortality could account for a vast array of behaviours not associated with death, or that a subtle death reminder is sufficient to motivate complex behaviour through unconscious processes. Despite our original intention to support and further explore the nature of MS effects, we found no empirical support for this hypothesis in the present study.

Acknowledgements

We were assisted in the data collection by student researchers Anna Berglann, Ingrid Kleiven, Marte Lovise Aagaard-Nilsen, Silje Johansen, Daniel Gunstveit, Astrid Bergwitz, Isabella Ahlvin, Line Molander, and Frida Ölander.

References

- Acharya, U. R., Joseph, K. P., Kannathal, N., Lim, C. M., & Suri, J. S. (2006). Heart rate variability: a review. *Medical and biological engineering and computing*, 44(12), 1031-1051.
- Arndt, J. (1999). Searching for the terror in terror management: Mortality salience and physiological indices of arousal and affect.
- Arndt, J., Allen, J. J., & Greenberg, J. (2001). Traces of terror: Subliminal death primes and facial electromyographic indices of affect. *Motivation and Emotion*, 25(3), 253-277.
- Arndt, J., Greenberg, J., & Cook, A. (2002). Mortality salience and the spreading activation of worldview-relevant constructs: exploring the cognitive architecture of terror management. *Journal of Experimental Psychology: General*, 131(3), 307.
- Arndt, J., Greenberg, J., Solomon, S., Pyszczynski, T., & Simon, L. (1997). Suppression, accessibility of death-related thoughts, and cultural worldview defense: Exploring the psychodynamics of terror management. *Journal of personality and social psychology*, 73(1), 5.
- Becker, E. (1973). *The Denial of Death*. New York, NY: The Free Press.
- Buhrmester, M., Kwang, T., & Gosling, S. D. (2011). Amazon's Mechanical Turk: A new source of inexpensive, yet high-quality, data? *Perspectives on Psychological Science*, 6(1), 3-5.
- Burke, B. L., Kosloff, S., & Landau, M. J. (2013). Death goes to the polls: A meta-analysis of mortality salience effects on political attitudes. *Political Psychology*, 34(2), 183-200.
- Burke, B. L., Martens, A., & Faucher, E. H. (2010). Two decades of terror management theory: A meta-analysis of mortality salience research. *Personality and Social Psychology Review*, 14(2), 155-195.
- Castano, E., Yzerbyt, V., Paladino, M.-P., & Sacchi, S. (2002). I belong, therefore, I exist: Ingroup identification, ingroup entitativity, and ingroup bias. *Personality and Social Psychology Bulletin*, 28(2), 135-143.
- Delaney, J., & Brodie, D. (2000). Effects of short-term psychological stress on the time and frequency domains of heart-rate variability. *Perceptual and Motor Skills*, 91(2), 515-524.
- Doosje, B., Ellemers, N., & Spears, R. (1995). Perceived intragroup variability as a function of group status and identification. *Journal of Experimental Social Psychology*, 31(5), 410-436.
- Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G* Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior research methods*, 39(2), 175-191.
- Federico, C. M., Jost, J. T., Pierro, A., & Kruglanski, A. W. (2007). The need for closure and political attitudes: Final report for the ANES pilot. *ANES Pilot Study Report*.
- Gailliot, M. T., Schmeichel, B. J., & Baumeister, R. F. (2006). Self-regulatory processes defend against the threat of death: Effects of self-control depletion and trait self-control on thoughts and fears of dying. *Journal of personality and social psychology*, 91(1), 49.
- Greenberg, J., Arndt, J., Simon, L., Pyszczynski, T., & Solomon, S. (2000). Proximal and distal defenses in response to reminders of one's mortality: Evidence of a temporal sequence. *Personality and Social Psychology Bulletin*, 26(1), 91-99.

- Greenberg, J., Porteus, J., Simon, L., Pyszczynski, T., & Solomon, S. (1995). Evidence of a terror management function of cultural icons: The effects of mortality salience on the inappropriate use of cherished cultural symbols. *Personality and Social Psychology Bulletin*, 21(11), 1221-1228.
- Greenberg, J., Pyszczynski, T., & Solomon, S. (1986). The causes and consequences of a need for self-esteem: A terror management theory. In *Public self and private self* (pp. 189-212): Springer.
- Greenberg, J., Pyszczynski, T., Solomon, S., Simon, L., & Breus, M. (1994). Role of consciousness and accessibility of death-related thoughts in mortality salience effects. *Journal of personality and social psychology*, 67(4), 627.
- Greenberg, J., Simon, L., Pyszczynski, T., Solomon, S., & Chatel, D. (1992). Terror management and tolerance: Does mortality salience always intensify negative reactions to others who threaten one's worldview? *Journal of personality and social psychology*, 63(2), 212.
- Greenberg, J., Solomon, S., & Pyszczynski, T. (1997). Terror management theory of self-esteem and cultural worldviews: Empirical assessments and conceptual refinements. In *Advances in experimental social psychology* (Vol. 29, pp. 61-139): Elsevier.
- Griskevicius, V., Tybur, J. M., Delton, A. W., & Robertson, T. E. (2011). The influence of mortality and socioeconomic status on risk and delayed rewards: a life history theory approach. *Journal of personality and social psychology*, 100(6), 1015.
- Hart, J. (2014). Toward an Integrative Theory of Psychological Defense. *Perspectives on Psychological Science*, 9(1), 19-39. doi:10.1177/1745691613506018
- Hauser, D. J., & Schwarz, N. (2016). Attentive Turkers: MTurk participants perform better on online attention checks than do subject pool participants. *Behavior research methods*, 48(1), 400-407.
- Hayes, J., & Schimel, J. (2018). Unintended effects of measuring implicit processes: The case of death-thought accessibility in mortality salience studies. *Journal of Experimental Social Psychology*, 74, 257-269.
- Hayes, J., Schimel, J., Arndt, J., & Faucher, E. H. (2010). A theoretical and empirical review of the death-thought accessibility concept in terror management research. *Psychological Bulletin*, 136(5), 699.
- Hayes, J., Schimel, J., Faucher, E. H., & Williams, T. J. (2008). Evidence for the DTA hypothesis II: Threatening self-esteem increases death-thought accessibility. *Journal of Experimental Social Psychology*, 44(3), 600-613.
- Heine, S. J., Harihara, M., & Niiya, Y. (2002). Terror management in Japan. *Asian Journal of Social Psychology*, 5(3), 187-196.
- Henry, E. A., Bartholow, B. D., & Arndt, J. (2010). Death on the brain: effects of mortality salience on the neural correlates of ingroup and outgroup categorization. *Social cognitive and affective neuroscience*, 5(1), 77-87.
- John, L. K., Loewenstein, G., & Prelec, D. (2012). Measuring the prevalence of questionable research practices with incentives for truth telling. *Psychological Science*, 0956797611430953.
- Jonas, E., Schimel, J., Greenberg, J., & Pyszczynski, T. (2002). The Scrooge effect: Evidence that mortality salience increases prosocial attitudes and behavior. *Personality and Social Psychology Bulletin*, 28(10), 1342-1353.
- Juhl, J., & Routledge, C. (2010). Structured terror: Further exploring the effects of mortality salience and personal need for structure on worldview defense. *Journal of personality*, 78(3), 969-990.
- Kees, J., Berry, C., Burton, S., & Sheehan, K. (2017). An analysis of data quality: Professional panels, student subject pools, and Amazon's Mechanical Turk. *Journal of Advertising*, 46(1), 141-155.
- 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. Retrieved from <https://psyarxiv.com/vef2c>
- Lane, R. D., McRae, K., Reiman, E. M., Chen, K., Ahern, G. L., & Thayer, J. F. (2009). Neural correlates of heart rate variability during emotion. *NeuroImage*, 44(1), 213-222.
- Martin, L. L., & Van den Bos, K. (2014). Beyond terror: Towards a paradigm shift in the study of threat and culture. *European Review of Social Psychology*, 25(1), 32-70.
- Morewedge, C. K., & Kahneman, D. (2010). Associative processes in intuitive judgment. *Trends in Cognitive Sciences*, 14(10), 435-440.
- Munafò, M. R., Nosek, B. A., Bishop, D. V., Button, K. S., Chambers, C. D., du Sert, N. P., . . . Ioannidis, J. P. (2017). A manifesto for reproducible science. *Nature Human Behaviour*, 1, 0021.
- Nosek, B. A., & Errington, T. M. (2019). What is replication? <https://doi.org/10.31222/osf.io/ut4g6t>.
- Pepper, G. V., Corby, D. H., Bamber, R., Smith, H., Wong, N., & Nettle, D. (2017). The influence of mortality and socioeconomic status on risk and delayed rewards: a replication with British participants. *PeerJ*, 5, e3580.
- Psychology Software Tools, I. (2012). E-Prime 2.0. Retrieved from <https://www.pstnet.com>
- Pyszczynski, T., Solomon, S., & Greenberg, J. (2015). Thirty years of terror management theory: From genesis to revelation. *Advances in experimental social psychology*, 52, 1-70.
- Roberts, J. R., & Maxfield, M. (2019). Mortality Salience and Age Effects on Charitable Donations. *American Behavioral Scientist*, 0002764219850864.
- Rodríguez-Ferreiro, J., Barberia, I., González-Guerra, J., & Vadillo, M. A. (2019). Are we truly special and unique? A replication of Goldenberg et al.(2001). *Royal Society open science*, 6(11), 191114.
- Rosenblatt, A., Greenberg, J., Solomon, S., Pyszczynski, T., & Lyon, D. (1989). Evidence for terror management theory: I. The effects of mortality salience on reactions to those who violate or uphold cultural values. *Journal of personality and social psychology*, 57(4), 681.
- Routledge, C., Ostafin, B., Juhl, J., Sedikides, C., Cathey, C., & Liao, J. (2010). Adjusting to death: the effects of mortality salience and self-esteem on psychological well-being, growth motivation, and maladaptive behavior. *Journal of personality and social psychology*, 99(6), 897.
- Schuler, E. R., Mlynski, C., & Wright, R. A. (2017). Influence of mortality salience on effort-related cardiovascular response to an identity-relevant challenge. *Motivation Science*, 3(2), 164-171.
- Silveira, S., Graupmann, V., Agthe, M., Gutyrchik, E., Blautzik, J., Demirçapa, I., . . . Reiser, M. (2013). Existential neuroscience: effects of mortality salience on the neurocognitive processing of attractive opposite-sex faces. *Social cognitive and affective neuroscience*, 9(10), 1601-1607.
- Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological science*, 22(11), 1359-1366.
- Simon, L., Greenberg, J., Harmon-Jones, E., Solomon, S., Pyszczynski, T., Arndt, J., & Abend, T. (1997). Terror management and cognitive-experiential self-theory: evidence that terror management occurs in the experiential system. *Journal of personality and social psychology*, 72(5), 1132.
- Simonsohn, U. (2015). Small telescopes: Detectability and the evaluation of replication results. *Psychological Science*, 26(5), 559-569.
- Singer, P. (2011). *The expanding circle: Ethics, evolution, and moral progress*: Princeton University Press.
- Stets, J. E. (2006). Identity Theory. In P. J. Burke (Ed.), *Contemporary social psychological theories* (pp. 88-110): Stanford University Press.
- The jamovi project. (2019). jamovi. (Version 0.9) Retrieved from <https://www.jamovi.org>
- Trafimow, D., & Hughes, J. S. (2012). Testing the death thought suppression and rebound hypothesis: Death thought accessibility following mortality salience decreases during a delay. *Social Psychological and Personality Science*, 3(5), 622-629.
- Turner, J. C., Oakes, P. J., Haslam, S. A., & McGarty, C. (1994). Self and collective: Cognition and social context. *Personality and Social Psychology Bulletin*, 20(5), 454-463.
- Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: the PANAS scales. *Journal of personality and social psychology*, 54(6), 1063.
- Yen, C.-L., & Cheng, C.-P. (2013). Researcher effects on mortality salience research: A meta-analytic moderator analysis. *Death Studies*, 37(7). doi:10.1080/07481187.2012.682290
- Zaleskiewicz, T., Gasiorowska, A., & Kesebir, P. (2015). The Scrooge effect revisited: Mortality salience increases the satisfaction derived from prosocial behavior. *Journal of Experimental Social Psychology*, 59, 67-76.