Sleeping over moral dilemmas modulates utilitarian decision-making

Nicola Cellini [1,2,3,4] · Marco Mercurio [1] · Michela Sarlo [5]

Abstract

Moral decision-making depends on the interaction between emotional and cognitive control processes, which are also affected by sleep. Here we aimed to assess the potential role of sleep in the modulation of moral decisions over time by testing the change in behavioral responses to moral dilemmas over time (1 week). Thirty-five young adults were tested twice, with one week between the sessions. In each session, participants were presented with 24 sacrificial (12 Footbridge- and 12 Trolley-type) and 6 everyday-type moral dilemmas. In sacrificial dilemmas, participants had to choose whether or not to kill one person to save more people (utilitarian choice), to judge how morally acceptable the proposed solution was, and how they felt in terms of valence and arousal during the decision. In everyday-type dilemmas, they had to decide whether to pursuit moral violations involving dishonest behavior. Between the sessions, the participants' sleep pattern was assessed via actigraphy. We observed that participants reduced the utilitarian choices in the second session, and this effect was more pronounced for the Trolley-type dilemmas. We also showed that after a week participants judged the utilitarian choices as less morally acceptable, but there was no change in self-reported emotional reactivity (i.e., valence, and arousal). Moreover, sleep efficiency was mildly negatively associated with the changes in decision choices and moral acceptability for the Footbridge-type dilemmas. Taken together, our data suggest that dealing with a moral situation engages several interacting factors that seem to go beyond the competing roles of cognitive and emotional processes.

Keywords Footbridge and Trolley dilemmas · Moral decision-making · Sleep · Stability · Emotion

✉ Nicola Cellini
nicola.cellini@unipd.it; cellini.nicola@gmail.com

Marco Mercurio
marcomerc93@gmail.com

Michela Sarlo
michela.sarlo@uniurb.it

1   Department of General Psychology, University of Padova, Via
    Venezia 8, 35131 Padova, Italy

2   Department of Biomedical Sciences, University of Padova, Via Ugo
    Bassi 58/B, 35131 Padova, Italy

3   Padova Neuroscience Center, University of Padova, Via Giuseppe
    Orus 2, 35131 Padova, Italy

4   Human Inspired Technology Center, University of Padova, Via
    Luzzatti 4, 35121 Padova, Italy

5   Department of Communication Sciences, Humanities and
    International Studies, University of Urbino Carlo Bo, Via Saffi 15,
    61029 Urbino, Italy

Introduction

In the last decades, several studies have highlighted the critical role of emotion in shaping decision-making, including moral judgments and actions (Lerner et al., 2015). A particular type of decision-making occurs when individuals are facing a moral dilemma, which is a situation where the outcomes of any decision are undesirable and the consequences of each choice are difficult to bear (Braunack-Mayer, 2001; Sinnott-Armstrong, 1987). The Trolley and the Footbridge dilemmas are prototypical examples of moral dilemmas. In the Trolley dilemma, five workmen are going to be killed by a runaway trolley and can be saved only by pulling a lever to redirect the trolley to a sidetrack, where a single workman will be killed. In the Footbridge dilemma, the five workmen can be saved only by pushing a large man down a bridge on the track in order to stop the trolley and save the five men. In both dilemmas, individuals are forced to make a decision: sacrificing one man to save five people, or letting the trolley kill the five men. Despite in both situations the cost/benefit ratio is exactly the same (i.e., 1/5), most people judge that pulling the lever in the Trolley dilemma is more morally acceptable than

pushing the man as in the Footbridge dilemma (Hauser et al., 2007; Lotto et al., 2014; Sarlo et al., 2012; Thomson, 1985).

According to the dual-process theory of moral judgment (Greene et al., 2004; Greene et al., 2001), the different decisions made in the Trolley and Footbridge dilemmas are due to an interaction between automatic emotional responses and rational cognitive control. Briefly, when facing a moral dilemma people immediately and automatically experience an aversive reaction. In the Footbridge dilemma, where the individuals are asked to push a man to die, this negative emotion is stronger (due to the direct action of killing someone) and leads to a fast rejection of the utilitarian resolution (i.e., an action based on the computation of the cost/benefit ratio). In the Trolley dilemma, the aversive reaction is supposed to be weaker since the killing occurs indirectly (i.e., pulling a lever) and the cognitive processes become dominant, leading to the utilitarian resolution (Greene et al., 2004; Greene et al., 2001). A corollary of this theory is that a reduction in the emotional drive is associated with increased cognitive control, which should lead to a higher propensity to endorse utilitarian resolutions. This idea has been supported by studies both on individuals with psychopathy (Pletti et al., 2016a; Tassy et al., 2013) and on patients with lesions to the ventromedial prefrontal cortex (vmPFC) (Ciaramelli et al., 2007; Koenigs et al., 2007), in which the integrity of the emotional processing system is compromised.

Therefore, based on the dual-process theory, it can be hypothesized that any intervention that can dampen the emotional activation elicited by dilemmas would increase the number of utilitarian resolutions. However, only a paucity of studies using moral dilemmas has manipulated the emotional state of the participants before decision-making (see Feinberg et al., 2012; Shapiro et al., 2012; Valdesolo & DeSteno, 2006). For example, Valdesolo and DeSteno (2006) induced positive or neutral affect in their participants using short comedy or documentary clips, before asking them to solve a Footbridge and a Trolley dilemma. They reported that participants who watched the funny clip tended to report more utilitarian responses than the ones watching the neutral videos. An alternative approach to videoclips, mindfulness, or reappraisal strategies for manipulating emotional processing is to capitalize on spontaneous "emotional regulation" intervention, such as sleeping. Indeed, according to the *Sleep to Forget, Sleep to Remember* (SFSR) hypothesis (Walker & van Der Helm, 2009), the specific neurobiological changes occurring during the rapid eye movement (REM) sleep (i.e., suppression of adrenergic activity, activation of amygdala-vmPFC-hippocampal networks, deactivation of the dorsolateral prefrontal cortex) are implicated in the reduction of the affective tone of recently acquired experiences while they are consolidated into long-term memory, thus leading to an emotional downregulation. Studies that have tested this model have reported mixed results, with several studies not showing any

reduction in the affecting tone after a period of sleep (Cellini et al., 2019; Cellini et al., 2016; Genzel et al., 2015; Tempesta et al., 2018). Moreover, as reported by two recent meta-analyses, the beneficial effect of sleep on the retention of emotional vs neutral information seems not to be supported (Lipinska et al., 2019; Schäfer et al., 2020).

Other models about the role of sleep in emotional regulation have been recently proposed. For example, the *emotional salience consolidation* model (ESC) proposes that physiological and subjective reactivity toward emotional stimuli is generally maintained (i.e., consolidated) or enhanced after a period of sleep (Baran et al., 2012; Bolinger et al., 2018; Bolinger et al., 2019; Pace-Schott et al., 2015; Sopp et al., 2017; Werner et al., 2015). This idea is grounded on the assumption that REM sleep can strengthen the association between a stimulus and its corresponding affective tone. An intermediate view between the SFSR and the ESC model has been proposed by Hutchison and Rathore (2015) and Genzel et al. (2015), who suggested that sleep facilitates the more adaptive emotional response for each specific event (e.g., fear enhancement to avoid a deadly situation, fear reduction for daily situations).

Based on the idea that (REM) sleep can promote spontaneous emotional regulation, i.e., by reducing the affective tone (as proposed by the SFSR model), and those strong emotional reactions would drive moral decision-making toward the rejection of utilitarian resolutions (as proposed by the dual-process theory), it might be hypothesized that sleeping after having being exposed to moral dilemma situations would change subsequent decision-making by increasing the probability to endorse utilitarian choices. However, the few studies investigating this issue showed that one session of either diurnal sleep or sleep deprivation does not affect moral decision-making (Cellini et al., 2017; Killgore et al., 2007; Olsen et al., 2010; D Tempesta et al., 2012). Indeed, Killgore et al. (2007) showed no effect of 53 h of sleep deprivation on the judgment of moral dilemmas. Similarly, Tempesta et al. (2012) showed no changes in moral judgment after a full night of sleep deprivation. More recently, another study showed no effect of a daytime nap on decision-making and emotional experience in moral dilemmas, although it was observed a negative association between theta activity during REM and increased self-rated unpleasantness during moral decisions (Cellini et al., 2017). The latter result is interesting since all the models previously described on sleep and emotions postulates that a relatively higher amount of REM sleep and/or several sleep cycles (i.e., several nights) might be needed to successfully implement emotion regulation through sleep. However, all the above studies have investigated only the acute effect of sleep and sleep deprivation on moral decision-making and judgment. Instead, as far as we know, no study has yet tested the effects of *several nights of sleep* on the resolution of moral dilemmas and the respective emotional reactions.

To fill this gap, in the current research we investigated the impact of one week of sleep on moral decision-making. Specifically, we tested participants twice, with one week between the two sessions, by using standardized moral dilemmas and assessing sleep using objective measures (i.e., actigraphy). Based on the SFSR model (Walker & van Der Helm, 2009), we hypothesized that in the second session, after one week from the initial exposure to a set of moral dilemmas and after several NREM-REM sleep cycles, participants would show a decrease in self-reported unpleasantness and emotional arousal. As predicted by the dual-process theory (Greene et al., 2004; Greene et al., 2001), the decrease in the affective tone due to several REM cycles would lead to an increase in the number of utilitarian choices as compared to the first session, with stronger effects expected for Footbridge- than for Trolley-type dilemmas. Therefore, a positive result would indicate that moral decision-making is indeed affected by natural sleep, but a sleep-induced modulation may require several nights to be effective, in line with the SFSR hypothesis.

Materials and Methods

Participants

Thirty-five University students (15 females) between the age of 19 and 29 years (mean age = 23.83 years, SD = 2.43) participated in this study. All participants were native Italian speakers and they were enrolled through advertisements posted at the University and they underwent an online screening to ensure they had no history of depression or sleep disorders. The study was approved by the local Ethics Committee and was in line with the Declaration of Helsinki. All participants gave written consent before participation.

Self-Reported Questionnaires

Self-reported questionnaires were employed to characterize the sleep quality, circadian preferences, and the depressive and anxiety level of the experimental sample.

Pittsburg Sleep Quality Index (PSQI)

The Italian version of the Pittsburg Sleep Quality Index (PSQI) was used to assess the level of self-reported sleep disturbances (Curcio et al., 2013). The questionnaire includes 19 items; the scores range from 0 and 21, with 0 indicating no sleep difficulties and 21 severe sleep difficulties (Buysse et al., 1989; Curcio et al., 2013; Mollayeva et al., 2016). The instrument has a good internal consistency, with a Cronbach's α of .835 for the Italian version, and .83 for the original version (Buysse et al., 1989). We excluded from the study participants

with a PSQI score higher than 10, which indicates participants at high risk of insomnia (Smith & Wegener, 2003).

Beck Depression Inventory-II (BDI-II)

To assess the severity of depressive symptomatology we employed the Italian version of the BDI-II (Ghisi et al., 2006). The BDI-II is composed of 21-item and the total score ranges from 0 to 63 with higher scores indicating greater depressive severity. The instrument has a good internal consistency, with a Cronbach's α of .87 for the Italian version, and .92–.94, depending on the sample, for the original version (Beck et al., 1996). We excluded from the study participants with a BDI-II score higher than 13, the cut-off for mild depressive symptomatology (Ghisi et al., 2006).

State-Trait Anxiety Inventory Y2 (STAI-Y2)

The STAI-Y2 is a self-report questionnaire composed of 20-items. The total score ranges from 20 to 100. Higher scores indicate greater anxiety levels (Spielberger, 2010). The original questionnaire had a Cronbach's α of .90 (Spielberger et al., 1983). Here we used the Italian version of the STAI-Y2 (Pedrabissi & Santinello, 1989), which has a Cronbach's α of .85–.90 depending on the sample.

Circadian Preferences

Circadian preferences were assessed using the reduced version of the Morningness–Eveningness Questionnaire (MEQr, Adan & Almirall, 1991; Natale et al., 2006a, 2006b). Using 5 items, with scores ranging from 4 to 35, this questionnaire categorizes participants into the evening (scores <11), intermediate (scores between 11 and 18), and morning types (scores >18). The Italian version of the questionnaire was used in the current study (Natale, 1999). The Cronbach's α of the MEQr is .83 for the Italian version and .84 for the original Spanish version (Adan & Natale, 2002).

Actigraphic Recording

To assess participants' sleep patterns, we asked them to wear the Actiwatch-64 (AW-64; Phillips Respironics, Portland, OR, US) for 7 days. This device is a reliable actigraph that estimates sleep parameters based on the level of movement activity (Cellini et al., 2013). Actigraphic data were collected for 7 days in 1-min epochs. Participants wore the actigraph on the non-dominant wrist and were instructed to press the AW-64 marker button every time they switched off/on the light to sleep and to get up from the bed, or when they had to remove the AW-64 for any reason (e.g., coming in contact with water). Actigraphic data were analyzed using the Actiware 6.2 software (Phillips Respironics, Portland, OR,

US), using the Medium threshold setting (40 activity counts/ epoch to define a waking state). In our analyses, we focused on the period confined between lights-off and lights-on. This interval was defined by combining actigraphic markers and sleep diary information. The following sleep parameters were extracted: the total sleep time (TST, min), as the number of minutes scored as sleep between lights off and lights on; sleep onset latency (SOL, min), the number of minutes between lights off and the first epoch scored as sleep; wake after sleep onset (WASO, min), the number of minutes scored as wake after sleep onset; and sleep efficiency (SE, %), the ratio between TST and total time spent in bed.

Experimental Task

We employed a moral decision-making task consisting in the resolution of 60 hypothetical moral problems selected from the standardized set from Lotto et al. (2014) and divided into two sets (A and B) of 30 dilemmas each. Twenty-four dilemmas were *sacrificial moral dilemmas*, in which the agent has to decide whether or not to sacrifice one individual to save more people; the remaining 6 were *everyday moral conflict situations* in which the agent has to decide whether or not to violate a moral obligation for personal advantages. The sacrificial dilemmas included 12 *Footbridge-type dilemmas*, describing sacrificing one individual as an intended means to save others, and 12 *Trolley-type dilemmas*, describing sacrificing one individual as a foreseen but unintended consequence of saving others (Fig. 1a).

Each dilemma was presented on two screens (Fig. 1b). The first one described the scenario, while the second one described a hypothetical action that the agent could perform (i.e., a utilitarian resolution for sacrificial dilemmas and a moral violation for everyday situations). In this second screen, participants had to choose whether or not they would perform the proposed action by pressing the buttons "Yes" or "No" on the keyboard (decision choice). After each dilemma, participants rated how they felt during the decision in terms of arousal (i.e., state of emotional activation) and valence (i.e., state of pleasantness), using a computerized version of the 9-point scales (1 to 9) of the Self-Assessment Manikin (Lang et al., 2008). Then, participants rated, on a scale from 0 (not at all) to 7 (completely), the extent to which the action was morally acceptable (Moral Acceptability), regardless of whether they decided to perform it or not. The task lasted about 30 min.

Participants performed the task in two experimental sessions (Fig. 1c), separated by 7 days. In each session, they were exposed to different dilemmas (set A or B), which were comparable for numerical consequences (i.e., the number of people to save or let die) and normative arousal and valence ratings (see Lotto et al., 2014). The order of the sets' presentation was counterbalanced between participants. Stimuli

presentation and data collection were performed using E-Prime 2.0 (Psychology Software Tools, Inc., Pittsburgh, PA, USA).

Experimental Procedure

Before the experimental sessions, all participants completed an online battery of questionnaires including the BDI-II, the PSQI, the STAI-Y2, and the MEQr. Then, they were scheduled for the two experimental sessions (Fig. 2c). When arrived at the laboratory, participants signed the informed consent and then completed two questionnaires: the Samn – Perelli Scale (Samn & Perelli, 1982) and the Stanford Sleepiness Scale (SSS; Hoddes et al., 1973) to evaluate fatigue and sleepiness levels, respectively. Then, participants received instructions for the task and completed three practice trials before the experiment began. One week later participants came back to the lab to perform the second experimental session, which had an identical procedure, but a different comparable set of dilemmas was used. The two sessions took place at the same time of the day to avoid any circadian influence on participants' responses.

During the seven days between the two sessions, participants were asked to complete a sleep diary and to wear an actigraph to assess their sleep patterns.

Statistical Analysis

Separate statistical analyses were performed for *sacrificial dilemmas* and *everyday moral situations*, as they are not directly comparable.

In each testing session, decision choices were computed for each participant as the proportion of utilitarian choices over the total number of response choices for each dilemma type (n = 12 for Trolley- and Footbridge-type and n = 6 for everyday moral situations). Response times were computed as the time from the onset of the hypothetical action to the decision choice (i.e., pressing the button "Yes" or "No"). Since response times showed a skewed distribution, they were log-transformed (natural logarithm) before the analyses. Mean valence and arousal ratings, as well as mean moral acceptability scores, were also computed separately for each participant and dilemma type. On each of our dependent variables (decision choices, response times, arousal, valence, and moral acceptability ratings) we conducted a 2 × 2 repeated measures analysis of variance (ANOVA) with *Session* (1 and 2) and *Dilemma Type* (Trolley-type and Footbridge-type) as within-subject factors. For the *everyday moral conflict situations,* we conducted an ANOVA with Session as the only within-subjects factor. For all analyses, partial eta squared (η_p^2) was reported as an estimate of effect size, and the Tukey HSD test was used for post-hoc comparisons.
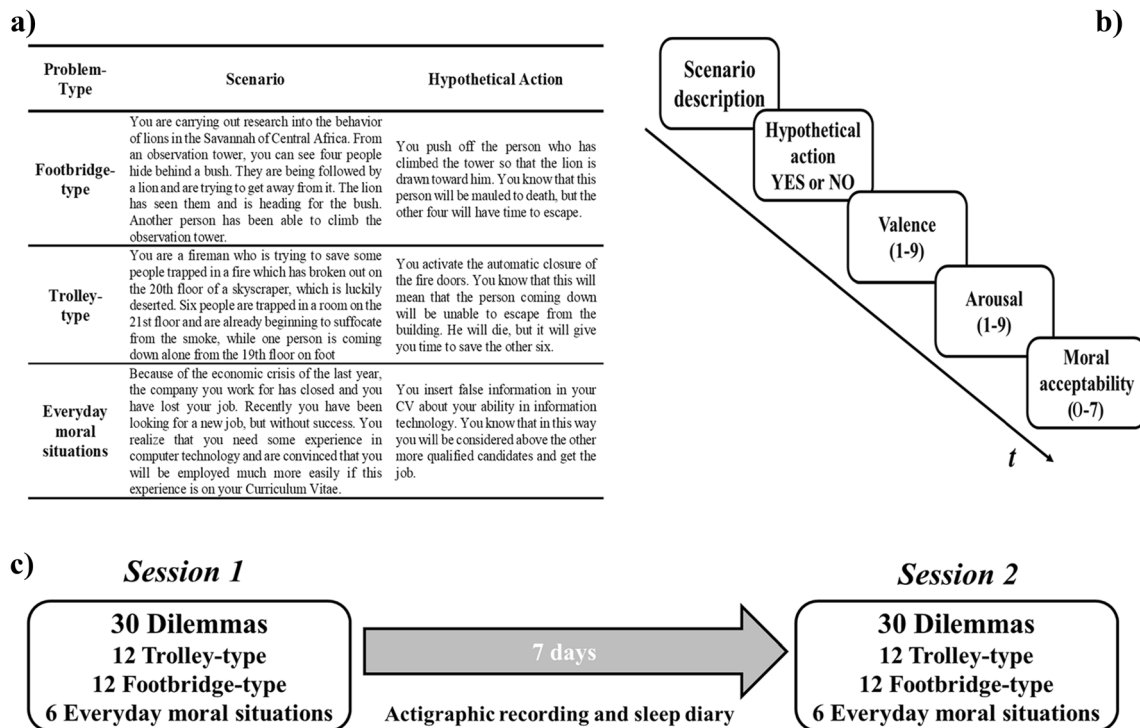
Fig. 1 **a** Examples of Footbridge- and Trolley-type dilemmas and everyday moral situations. **b** Schematic representation of the experimental task. After 3 practice dilemmas, participants were exposed to 30 dilemmas (12 Trolley-type, 12 Footbridge-type, 6 everyday moral situations). Each dilemma started with a first screen where the scenario was described, followed by a second screen where a hypothetical action was proposed: here the participants had to decide whether or not to perform it. After that, participants rated how they felt during the decision in terms of valence (1–9 scores) and arousal (1–9 scores). Then they rated to what extent the proposed action was morally acceptable (0–7 scores). **c** Schematic representation of experimental procedure. Participants performed two experimental sessions, each composed of 30 different dilemmas, and separated by 7 days, during which the participants' sleep was monitored through actigraphy and sleep diaries
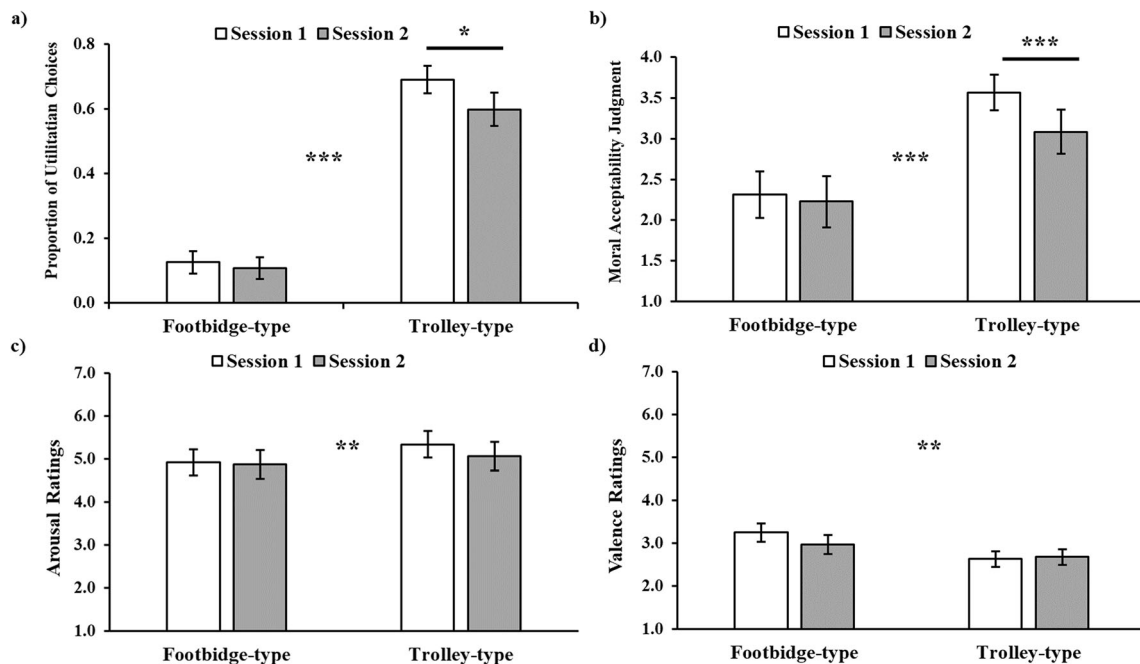

Fig. 2 **a** Proportions of utilitarian choices **b** Ratings of moral acceptability **c** Arousal ratings **d** Valence ratings as a function of Dilemma Type (Trolley-type and Footbridge-type) and Session in sacrificial moral dilemmas. Error bars represent standard error. ***: $p < .001$, *: $p < .05$

To assess the stability of individuals' decisions and ratings between the two sessions, we computed the intraclass correlation coefficients (ICCs) for each behavioral variable (decision choices, response times, arousal ratings, valence ratings, and moral acceptability ratings) separately for the three types of dilemmas (Footbridge-type, Trolley-type, Everyday situations). ICCs are a commonly used technique to assess the relative reliability between measurements (Atkinson & Nevill, 1998). We compute ICCs using two-way random effects, absolute agreement, and single measurement (ICCs (2,1)). The ICCs (2,1) accounts for both the agreement of performance between the two sessions in the same individual (within-subject change) and the change in the mean performance of participants as a group between the sessions (i.e., systematic change in mean). We considered ICCs below .50 as poor agreement, between .50 and .75 as moderate agreement, between .75 to .90 as good agreement, and above .90 excellent agreement (Koo & Li, 2016).

Lastly, Pearson's correlations were conducted to explore the associations between sleep quality over the week (as indexed by the sleep efficiency) and the change in behavioral responses on the moral dilemma tasks between the two sessions (computed as Session 2 scores minus Session 1 scores). For all the analyses, a p value < .05 was considered statistically significant.

Results

The descriptive statistics of the sample are presented in Table 1.

Sacrificial Moral Dilemmas

Means and standard deviations for the behavioral responses to the sacrificial moral dilemmas are reported in Table 2.

Decision Choices

The ANOVA on the proportion of utilitarian choices showed a significant *Dilemma Type* main effect ($F_{1,34} = 175.64$, $p < .001$, $\eta_p^2 = .84$), with a higher proportion of utilitarian choices (i.e., sacrificing one individual to save more people) for the Trolley- as compared to the Footbridge-type dilemmas. We also observed a *Session* main effect ($F_{1,34} = 6.71$, $p = .014$, $\eta_p^2 = .16$), showing a reduction in the proportion of utilitarian choices in the second session. The *Dilemma Type×Session* interaction did not reach significance ($F_{1,34} = 2.89$, $p = .098$, $\eta_p^2 = .08$, Fig. 2a). However, it can be observed that the number of utilitarian choices decreased from the first to the second session especially for the Trolley-type dilemmas ($p = .027$).

Table 1 Demographics, questionnaire measures, and sleep parameters of the sample

	Mean	SD
Demographics and baseline measures		
Age	23.82	2.42
Gender (F/M)	15/20	
BDI-II	6.23	3.65
STAI-Y2	40.86	7.81
PSQI	5.06	2.11
MEQr	12.94	4.20
Sleep Parameters		
Bed Time (hh:mm)	01:07	01:19
Wake Time (hh:mm)	08:45	01:08
Time in Bed (min)	458.19	36.73
Sleep Latency (min)	7.51	4.65
WASO (min)	49.01	13.26
Total Sleep Time (min)	401.67	36.97
Sleep Efficiency (%)	87.65	2.65

BDI-II Beck Depression Inventory-II, *STAI-Y2* State-Trait Anxiety Inventory Y2, *PSQI* Pittsburg Sleep Quality Index, *MEQr* Morningness-Eveningness Questionnaire reduced version, *WASO* Wake After Sleep Onset

Response Times

A significant *Dilemma Type* main effect was found ($F_{1,34} = 81.52$, $p < .001$, $\eta_p^2 = .71$), with faster responses for the Footbridge- than for the Trolley-type dilemmas. The analysis also revealed a *Session* main effect ($F_{1,34} = 39.02$, $p < .001$, $\eta_p^2 = .53$), with overall faster responses in the second session as compared to the first one. The *Dilemma Type×Session* interaction was not significant ($F_{1,34} = 0.18$, $p = .68$, $\eta_p^2 = .01$; Fig. S1).

Moral Acceptability

A significant *Dilemma Type* main effect was observed ($F_{1,34} = 22.92$, $p < .001$, $\eta_p^2 = .40$), with utilitarian resolutions being rated as more morally acceptable for Trolley-type dilemmas. We also observed a significant *Session* main effect ($F_{1,34} = 6.47$, $p = .015$, $\eta_p^2 = .16$), which can be explained by the significant *Dilemma Type×Session* interaction ($F_{1,34} = 4.73$, $p = .037$, $\eta_p^2 = .12$; Fig. 2b), showing that Trolley-type dilemmas were rated as less morally acceptable in the second session compared to the first one ($p = .004$).

Arousal and Valence Ratings

The ANOVA on arousal ratings showed a significant *Dilemma Type* main effect ($F_{1,34} = 8.71$, $p = .006$, $\eta_p^2 = .20$), with higher arousal ratings for the Trolley-type dilemmas. No

Table 2 Means ± standard deviations of all the dependent variables for the sacrificial moral dilemmas in the two sessions as a function of dilemma type

	Session 1		Session 2	
	Footbridge-Type	Trolley-Type	Footbridge-Type	Trolley-Type
Decision choice	0.13 ± 0.21	0.69 ± 0.25	0.11 ± 0.20	0.60 ± 0.31
Response times (ln ms)	9.20 ± 0.34	9.61 ± 0.42	8.96 ± 0.29	9.40 ± 0.42
Moral acceptability	2.31 ± 1.68	3.46 ± 1.30	2.23 ± 1.84	3.08 ± 1.61
Arousal ratings	4.93 ± 1.78	5.34 ± 1.86	4.88 ± 1.96	5.06 ± 1.94
Valence ratings	3.25 ± 1.29	2.63 ± 1.11	2.97 ± 1.28	2.68 ± 1.08

ln: natural logarithm

significant changes were observed for the *Session* main effect ($F_{1,34} = 2.35$, $p = .13$, $\eta_p^2 = .06$) or the *Dilemma Type × Session* interaction ($F_{1,34} = 1.75$, $p = .19$, $\eta_p^2 = .05$; Fig. 2c).

The analysis of valence ratings showed again a significant *Dilemma Type* main effect ($F_{1,34} = 10.99$, $p = .002$, $\eta_p^2 = .24$), with lower valence ratings for the Trolley-type dilemmas. No significant changes were observed for the *Session* main effect ($F_{1,34} = 0.97$, $p = .33$, $\eta_p^2 = .06$) or for the *Dilemma Type × Session* interaction ($F_{1,34} = 3.81$, $p = .059$, $\eta_p^2 = .10$, Fig. 2d).

Everyday Moral Situations

Means and standard deviations for the behavioral responses to everyday moral situations are reported in Table S3. The ANOVAs on decision choices, moral acceptability, and affecting ratings (i.e., valence and arousal) did not show any significant *Session* main effect (all $Fs_{1,34} \leq 0.90$, $ps \geq .34$, $\eta_p^2s \leq .02$). The only significant effect was observed for decision times, with faster response times in the second session as compared to the first one ($F_{1,34} = 12.31$, $p = .001$, $\eta_p^2 = .27$).

Exploratory Correlational Analyses

When exploring the association between sleep quality over the week (as indexed by the sleep efficiency, SE) and the changes in behavioral variables of the moral dilemma task between the two sessions, we observed a negative association between SE and the changes in decision choices and moral acceptability for the Footbridge-type dilemmas only ($r = -.47$, $p = .005$ and $r = -.31$, $p = .073$, respectively; Fig. 3), with the latter result not reaching significance. This indicates that, when confronted with the Footbridge-type dilemmas, participants with greater SE reduced their utilitarian choices and considered the utilitarian resolutions as less morally acceptable in the second as compared to the first session. No significant associations were observed between other task-related variables and SE (all $ps > .17$, all $rs < |.23|$).

Discussion

In the current study, we investigated the impact of several nights of sleep (i.e., one week) on moral decision-making. Based on the dual-process theory of moral judgment (Greene et al., 2004; Greene et al., 2001), the type of resolution of a moral dilemma depends on a competition between the level of the emotional activation and the allocation of cognitive resources. Therefore, reducing the emotional reaction elicited by a moral situation should allow cognitive processes to become predominant, thus increasing the number of utilitarian choices (i.e., sacrificing one person to save more people). Based on the SFSR model (Walker & van Der Helm, 2009), which proposes that sleep benefits the consolidation of declarative aspects of experiences and facilitates the processing of the related emotional information while reducing its affective tone, we hypothesized that, during the seven nights (and the respective several NREM-REM sleep cycles) between the first and second exposure to a set of moral dilemmas, the spontaneous emotion regulation induced by naturally-occurring REM sleep would dampen the emotional activation elicited by dilemmas and, consequently, increase cognitive control favoring rational resolutions based on the cost/benefit ratio. Therefore, we expected an increase in the number of utilitarian choices and a decrease in the experienced unpleasantness and arousal when participants were exposed to the moral dilemmas during the second as compared to the first session.

However, our results were in contrast to our hypotheses. Indeed, in the second session, we found no significant changes in valence or arousal ratings reported during the resolution of the dilemmas, suggesting that sleep was not able to modulate the subjective emotional responses to the dilemmas, which would be in contrast with the SFSR hypothesis (Walker & van Der Helm, 2009). In particular, we observed a reliable decrease in the proportion of utilitarian resolutions, especially for Trolley-type dilemmas, for which a clear-cut reduction in the judgments of moral acceptability was also obtained. Indeed, the processing of moral dilemmas over one week reduced the utilitarian inclinations based on cost/benefit analysis, i.e., participants were less willing to perform harmful actions, and
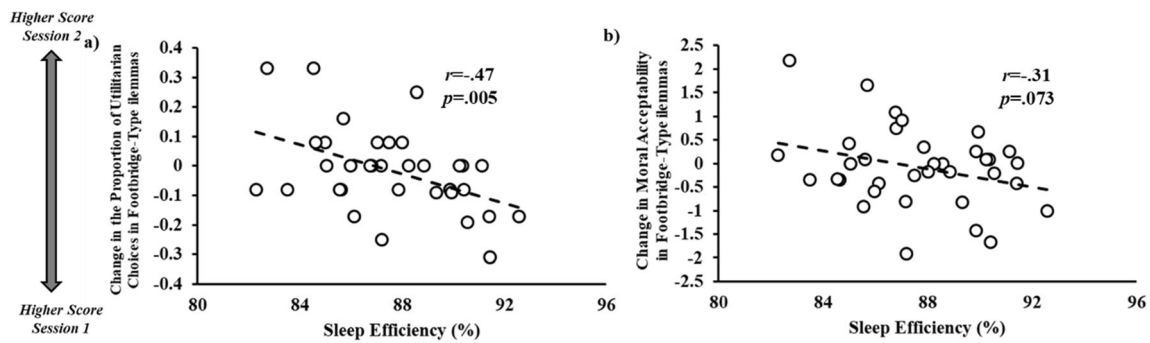
Fig. 3 a Changes in the proportion of utilitarian choices for Footbridge-type dilemmas as a function of sleep efficiency across the week. b Changes in moral acceptability scores for Footbridge-type dilemmas as a function of sleep efficiency across the week

judged these actions as less morally acceptable even when aimed at the greater good. This effect was more pronounced for the Trolley-type dilemmas, as for Footbridge-type dilemmas the initial proportion of utilitarian choices and ratings of moral acceptability were already very low (i.e., 0.13 and 2.31, respectively; see Fig. 2a). The observed changes in decision choices and moral judgments might be due either to the daytime reprocessing of the dilemmatic situations or to the effects of the several sleep cycles, or both. Unfortunately, the employed experimental design does not allow to disentangle the underlying causal relationships. However, in the case of Footbridge-type dilemmas, we observed a negative relationship between sleep efficiency and change in the proportion of utilitarian choices, i.e. the higher the quality of the restorative behavior of sleep during the week, the higher the reduction in utilitarian choices in the second as compared to the first session. A similar negative correlation (albeit non-significant) was observed with the judgments of moral acceptability. Although these results should be taken with caution, as the correlations were explorative and no multiple testing correction was applied, such sleep-related effects, as well as the finding of no change in subjective emotional reactivity across sessions, is at odds with the *SFSR* hypothesis (Walker & van Der Helm, 2009). Since we used actigraphy, rather than polysomnography, we can only speculate about the architecture of our participants' sleep. Nevertheless, we can suppose that higher sleep efficiency in our sample would be associated with a higher amount of REM sleep. Therefore, we should have expected that higher sleep efficiency would induce a reduction in the affective tone elicited by the dilemmatic situations, leading to an increase in the proportion of utilitarian choices, as predicted by the dual-process theory of moral judgment (Greene et al., 2004; Greene et al., 2001). Instead, we observed the opposite picture, suggesting that the better is someone's sleep, the less they judge the utilitarian resolution of the dilemmas as morally acceptable, and the less they are eager to opt for that resolution. These results are more in line with the *ESC* model, which proposes that sleep, by strengthening the association between an event and its corresponding emotional tone, facilitates the preservation of the affective tone of experiences (Baran et al., 2012;

Bolinger et al., 2018; Bolinger et al., 2019; Pace-Schott et al., 2011; Werner et al., 2015). As suggested by Tempesta et al. (2018), the maintenance of salience in critical situations such as a moral dilemma has a strong evolutionary value since it may facilitate both the memory of a significant emotional event and the degree of threat associated with it (e.g., aversive personal consequences; Sarlo et al., 2014). We can speculate that, in our study, the persistence of the emotional impact of the sacrificial dilemmas during the week might have preserved, and indeed emphasized, the representation of the moral norm against killing or harming, thus reducing the probability of choosing the utilitarian option. According to the rule-based approach (Nichols, 2002), representation of rules contribute to moral judgment as individuals can rely on a *normative theory*, which is a body of norms and rules describing what is allowed and what is not that would be acquired during the development. These rules are supposed to be stored in long-term memory (Bunge, 2004), and when an individual is facing a challenging situation such as the resolution of moral dilemmas, the final decision is the result of the interaction between the immediate emotional reactivity to the dilemmas and the stored representation of the rule that is challenged (i.e., *killing is always wrong*!; see Pletti et al., 2016b). On these bases, we can speculate that several nights of sleep after being exposed to sacrificial moral dilemmas may, on the one hand, preserve the emotional reactivity to the dilemmas and, on the other hand, reactivate and strengthen the representation of the moral rules motivating aversion to harming others, which would result in a lower number of utilitarian choices in similar situations.

It is worth noting that the responses to the sacrificial dilemmas were consistent in the two sessions across most of our participants, as highlighted by the ICCs analysis (i.e., agreements ranging from moderate to excellent, see supplemental material). This result indicates that the way each participant responds to the moral dilemmas is pretty stable even after a week and quite resistant to change over time, in line with the idea that moral decision-making is largely affected by a wide range of personality traits (e.g., Mudrack, 2006) as well as by individual differences in sensitivity to costs and benefits (Moore et al., 2011).

Overall, we showed for the first time that moral decision-making and judgment change over time, leading to a less utilitarian inclination, although the subjective experience to the dilemmas exposure (i.e., arousal and valence ratings) is preserved. Sleep might promote this change in moral decision-making, possibly by reactivating the representation of those moral rules that emotion seems to emphasize. However, further research is needed to test this hypothesis.

Consistently with previous literature, we also showed that participants provided less utilitarian choices for the Footbridge-type (more emotionally-driven) than the Trolley-type dilemmas (more cognitively-driven), as well as faster response times for the Footbridge-type dilemmas, indicating a more automatic and less conflicting decision processing, favoring the rejection of utilitarian resolutions (Cellini et al., 2017; Cushman et al., 2006; Hauser et al., 2007; Lotto et al., 2014; Moore et al., 2008; Sarlo et al., 2012).

At the same time, we showed for the first time that the proportion of utilitarian resolutions decreased over time, especially for Trolley-type dilemmas, for which a clear-cut reduction in the judgments of moral acceptability was also obtained. The latter result is in line with a recent study showing that participants judged the utilitarian choices as less morally acceptable after one daytime nap or a similar period of wakefulness (i.e., 2 h; Cellini et al., 2017). However, unlike the present study, Cellini et al. (2017) did not observe significant changes in the number of utilitarian choices. Taken together, these findings indicate that both moral judgments and decision choices tend to change over time, although with a different trajectory (e.g., within few hours for the moral judgment and in several days for decision choice). This result is interesting since the literature has often reported dissociations between moral judgments and behavioral choices, suggesting the existence of different underlying processes, with judgments relying more on normative prescriptions and beliefs (Nichols & Mallon, 2006), whereas choices of action are more affected by emotional processing and personal experiences (Loewenstein & Lerner, 2003; Tassy et al., 2012; Zeelenberg et al., 2008). Here, we showed that moral judgments and behavioral choices seem to converge over time.

The current results should be interpreted taking into account the study's limitations. First of all, this study was correlational, with a limited sample size ($N = 35$), and no independent manipulation of the passage of time and sleep cycles was performed. Second, we assessed sleep patterns over a week using actigraphy instead of polysomnography. Although actigraphy can be considered a reliable tool to assess sleep patterns (Cellini et al., 2013), it does not allow the assessment of sleep architecture and the proportion of the different sleep stages (e.g., REM sleep). Third, the sample was composed only of Italian young adults, therefore the results may not be generalized to other age groups or different cultures.

In conclusion, here we showed that after a week, when individuals were re-exposed to moral dilemmas, their moral decision-making became less utilitarian, with no changes in self-reported emotional reactivity. Moreover, it was only mildly modulated by their sleep parameters and showed stable intraindividual patterns of variability across sessions. Taken together, our data suggest that dealing with a moral situation engages several interacting factors that seem to go beyond the competing roles of cognitive and emotional processes.

Declaration

Conflict of Interest   No conflict of interest to disclose.

References

Adan, A., & Almirall, H. (1991). Horne & Östberg morningness-eveningness questionnaire: A reduced scale. *Personality and Individual Differences, 12*(3), 241–253.

Adan, A., & Natale, V. (2002). Gender differences in morningness–eveningness preference. *Chronobiology International, 19*(4), 709–720.

Atkinson, G., & Nevill, A. M. (1998). Statistical methods for assessing measurement error (reliability) in variables relevant to sports medicine. *Sports Medicine, 26*(4), 217–238.

Baran, B., Pace-Schott, E. F., Ericson, C., & Spencer, R. M. (2012). Processing of emotional reactivity and emotional memory over sleep. *Journal of Neuroscience, 32*(3), 1035–1042.

Beck, A. T., Steer, R. A., & Brown, G. K. (1996). Beck depression inventory-II. *San Antonio, 78*(2), 490–498.

Bolinger, E., Born, J., & Zinke, K. (2018). Sleep divergently affects cognitive and automatic emotional response in children. *Neuropsychologia, 117*, 84–91.

Bolinger, E., Cunningham, T. J., Payne, J. D., Bowman, M. A., Bulca, E., Born, J., & Zinke, K. (2019). Sleep's benefits to emotional processing emerge in the long term. *Cortex, 120*, 457–470.

Braunack-Mayer, A. J. (2001). What makes a problem an ethical problem? An empirical perspective on the nature of ethical problems in general practice. *Journal of Medical Ethics, 27*(2), 98–103.

Bunge, S. A. (2004). How we use rules to select actions: A review of evidence from cognitive neuroscience. *Cognitive, Affective, & Behavioral Neuroscience, 4*(4), 564–579.

Buysse, D. J., Reynolds III, C. F., Monk, T. H., Berman, S. R., & Kupfer, D. J. (1989). The Pittsburgh sleep quality index: A new instrument for psychiatric practice and research. *Psychiatry Research, 28*(2), 193–213. https://doi.org/10.1016/0165-1781(89)90047-4

Cellini, N., Buman, M. P., McDevitt, E. A., Ricker, A. A., & Mednick, S. C. (2013). Direct comparison of two actigraphy devices with polysomnographically recorded naps in healthy young adults. *Chronobiology International, 30*(5), 691–698. https://doi.org/10.3109/07420528.2013.782312

Cellini, N., Torre, J., Stegagno, L., & Sarlo, M. (2016). Sleep before and after learning promotes the consolidation of both neutral and emotional information regardless of REM presence. *Neurobiology of Learning and Memory, 133*, 136–144. https://doi.org/10.1016/j.nlm.2016.06.015

Cellini, N., Lotto, L., Pletti, C., & Sarlo, M. (2017). Daytime REM sleep affects emotional experience but not decision choices in moral dilemmas. *Scientific Reports, 7*(1), 11059.

Cellini, N., Mercurio, M., & Sarlo, M. (2019). The fate of emotional memories over a week: Does sleep play any role? *Frontiers in Psychology, 10*, 481.

Ciaramelli, E., Muccioli, M., Ladavas, E., & di Pellegrino, G. (2007). Selective deficit in personal moral judgment following damage to ventromedial prefrontal cortex. *Social Cognitive and Affective Neuroscience, 2*(2), 84–92.

Curcio, G., Tempesta, D., Scarlata, S., Marzano, C., Moroni, F., Rossini, P. M., et al. (2013). Validity of the Italian version of the Pittsburgh sleep quality index (PSQI). *Neurological Sciences, 34*(4), 511–519. https://doi.org/10.1007/s10072-012-1085-y

Cushman, F., Young, L., & Hauser, M. (2006). The role of conscious reasoning and intuition in moral judgment: Testing three principles of harm. *Psychological Science, 17*(12), 1082–1089.

Feinberg, M., Willer, R., Antonenko, O., & John, O. P. (2012). Liberating reason from the passions: Overriding intuitionist moral judgments through emotion reappraisal. *Psychological Science, 23*(7), 788–795.

Genzel, L., Spoormaker, V., Konrad, B., & Dresler, M. (2015). The role of rapid eye movement sleep for amygdala-related memory processing. *Neurobiology of Learning and Memory, 122*, 110–121.

Ghisi, M., Flebus, G., Montano, A., Sanavio, E., & Sica, C. (2006). *Beck depression inventory-II. Manuale.* Firenze: O.S,. Organizzazioni Speciali.

Greene, J. D., Sommerville, R. B., Nystrom, L. E., Darley, J. M., & Cohen, J. D. (2001). An fMRI investigation of emotional engagement in moral judgment. *Science, 293*(5537), 2105–2108.

Greene, J. D., Nystrom, L. E., Engell, A. D., Darley, J. M., & Cohen, J. D. (2004). The neural bases of cognitive conflict and control in moral judgment. *Neuron, 44*(2), 389–400.

Hauser, M., Cushman, F., Young, L., Kang-Xing Jin, R., & Mikhail, J. (2007). A dissociation between moral judgments and justifications. *Mind & Language, 22*(1), 1–21.

Hoddes, E., Zarcone, V., Smythe, H., Phillips, R., & Dement, W. (1973). Quantification of sleepiness: A new approach. *Psychophysiology, 10*(4), 431–436.

Hutchison, I. C., & Rathore, S. (2015). The role of REM sleep theta activity in emotional memory. *Frontiers in Psychology, 6*.

Killgore, W. D., Killgore, D. B., Day, L. M., Li, C., Kamimori, G. H., & Balkin, T. J. (2007). The effects of 53 hours of sleep deprivation on moral judgment. *Sleep, 30*(3), 345–352.

Koenigs, M., Young, L., Adolphs, R., Tranel, D., Cushman, F., Hauser, M., & Damasio, A. (2007). Damage to the prefrontal cortex increases utilitarian moral judgements. *Nature, 446*(7138), 908–911.

Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine, 15*(2), 155–163.

Lang, P. J., Bradley, M. M., & Cuthbert, B. N. (2008). *International affective picture system (IAPS): Affective ratings of pictures and instruction manual.* University of Florida.

Lerner, J. S., Li, Y., Valdesolo, P., & Kassam, K. S. (2015). Emotion and decision making. *Annual Review of Psychology, 66*, 799–823.

Lipinska, G., Stuart, B., Thomas, K. G., Baldwin, D. S., & Bolinger, E. (2019). Preferential consolidation of emotional memory during sleep: A meta-analysis. *Frontiers in Psychology, 10*.

Loewenstein, G., & Lerner, J. S. (2003). The role of affect in decision making. *Handbook of affective science, 619*(642), 3.

Lotto, L., Manfrinati, A., & Sarlo, M. (2014). A new set of moral dilemmas: Norms for moral acceptability, decision times, and emotional salience. *Journal of Behavioral Decision Making, 27*(1), 57–65.

Mollayeva, T., Thurairajah, P., Burton, K., Mollayeva, S., Shapiro, C. M., & Colantonio, A. (2016). The Pittsburgh sleep quality index as a screening tool for sleep dysfunction in clinical and non-clinical samples: A systematic review and meta-analysis. *Sleep Medicine Reviews, 25*, 52–73.

Moore, A. B., Clark, B. A., & Kane, M. J. (2008). Who shalt not kill? Individual differences in working memory capacity, executive control, and moral judgment. *Psychological Science, 19*(6), 549–557.

Moore, A. B., Stevens, J., & Conway, A. R. (2011). Individual differences in sensitivity to reward and punishment predict moral judgment. *Personality and Individual Differences, 50*(5), 621–625.

Mudrack, P. E. (2006). Moral reasoning and personality traits. *Psychological Reports, 98*(3), 689–698.

Natale, V. (1999). Validazione di una scala ridotta di mattutinità (r-MEQ) [Validation of a shortened morningness scale (r-MEQ)]. *Bollettino di Psicologia Applicata, 229*, 19–26.

Natale, V., Esposito, M. J., Martoni, M., & Fabbri, M. (2006a). Validity of the reduced version of the Morningness–Eveningness Questionnaire. *Sleep and Biological Rhythms, 4*(1), 72–74. https://doi.org/10.1111/j.1479-8425.2006.00192.x

Natale, V., Grandi, C. A., Fabbri, M., Tonetti, L., Martoni, M., & Esposito, M. J. (2006b). Additional validity evidence for the reduced version of the Morningness-Eveningness Questionnaire (MEQr). *Sleep and Hypnosis, 8*(2), 47–53.

Nichols, S. (2002). Norms with feeling: Towards a psychological account of moral judgment. *Cognition, 84*(2), 221–236.

Nichols, S., & Mallon, R. (2006). Moral dilemmas and moral rules. *Cognition, 100*(3), 530–542.

Olsen, O. K., Pallesen, S., & Eid, J. (2010). The impact of partial sleep deprivation on moral reasoning in military officers. *Sleep, 33*(8), 1086–1090.

Pace-Schott, E. F., Shepherd, E., Spencer, R. M., Marcello, M., Tucker, M., Propper, R. E., & Stickgold, R. (2011). Napping promotes inter-session habituation to emotional stimuli. *Neurobiology of Learning and Memory, 95*(1), 24–36.

Pace-Schott, E. F., Germain, A., & Milad, M. R. (2015). Effects of sleep on memory for conditioned fear and fear extinction. *Psychological Bulletin, 141*(4), 835–857. https://doi.org/10.1037/bul0000014

Pedrabissi, L., & Santinello, M. (1989). Verifica della validità dello STAI forma Y di Spielberger. *Giunti Organizzazioni Speciali.*

Pletti, C., Lotto, L., Buodo, G., & Sarlo, M. (2016a). It's immoral, but I'd do it! Psychopathy traits affect decision-making in sacrificial dilemmas and in everyday moral situations. British Journal of Psychology.

Pletti, C., Lotto, L., Tasso, A., & Sarlo, M. (2016b). Will I regret it? Anticipated negative emotions modulate choices in moral dilemmas. *Frontiers in Psychology, 7*, 1918.

Samn, S. W., & Perelli, L. P. (1982). *Estimating aircraft fatigue: A technique with application to airline operations.* Brooks AFB, Tex: USAF School of Medicine.

Sarlo, M., Lotto, L., Manfrinati, A., Rumiati, R., Gallicchio, G., & Palomba, D. (2012). Temporal dynamics of cognitive–emotional interplay in moral decision-making. *Journal of Cognitive Neuroscience, 24*(4), 1018–1029.

Sarlo, M., Lotto, L., Rumiati, R., & Palomba, D. (2014). If it makes you feel bad, don't do it! Egoistic rather than altruistic empathy modulates neural and behavioral responses in moral dilemmas. *Physiology and Behavior, 130*, 127–134.

Schäfer, S. K., Wirth, B. E., Staginnus, M., Becker, N., Michael, T., & Sopp, M. R. (2020). Sleep's impact on emotional recognition memory: A meta-analysis of whole-night, nap, and REM sleep effects. *Sleep Medicine Reviews, 101280.*

Shapiro, S. L., Jazaieri, H., & Goldin, P. R. (2012). Mindfulness-based stress reduction effects on moral reasoning and decision making. *The Journal of Positive Psychology, 7*(6), 504–515.

Sinnott-Armstrong, W. (1987). Moral realisms and moral dilemmas. *The Journal of Philosophy, 84*(5), 263–276.

Smith, M. T., & Wegener, S. T. (2003). Measures of sleep: The insomnia severity index, medical outcomes study (MOS) sleep scale, Pittsburgh sleep diary (PSD), and Pittsburgh sleep quality index (PSQI). *Arthritis Care & Research: Official Journal of the American College of Rheumatology, 49*(S5), S184–S196.

Sopp, M. R., Michael, T., Weeß, H.-G., & Mecklinger, A. (2017). Remembering specific features of emotional events across time: The role of REM sleep and prefrontal theta oscillations. *Cognitive, Affective, & Behavioral Neuroscience, 17*(6), 1186–1209.

Spielberger, C. D., Gorsuch, R. L., Lushene, R., Vagg, P. R., & Jacobs, G. A. (1983). *Manual for the state-trait anxiety inventory.* Consulting Psychologists Press.

Tassy, S., Oullier, O., Duclos, Y., Coulon, O., Mancini, J., Deruelle, C., Attarian, S., Felician, O., & Wicker, B. (2012). Disrupting the right prefrontal cortex alters moral judgement. *Social Cognitive and Affective Neuroscience, 7*, 282–288. https://doi.org/10.1093/scan/nsr008

Tassy, S., Deruelle, C., Mancini, J., Leistedt, S., & Wicker, B. (2013). High levels of psychopathic traits alters moral choice but not moral judgment. *Frontiers in Human Neuroscience, 7.*

Tempesta, D., Couyoumdjian, A., Moroni, F., Marzano, C., De Gennaro, L., & Ferrara, M. (2012). The impact of one night of sleep deprivation on moral judgments. *Social Neuroscience, 7*(3), 292–300.

Tempesta, D., Socci, V., De Gennaro, L., & Ferrara, M. (2018). Sleep and emotional processing. *Sleep Medicine Reviews, 40*, 183–195. https://doi.org/10.1016/j.smrv.2017.12.005

Thomson, J. J. (1985). The trolley problem. *The Yale Law Journal, 94*(6), 1395–1415.

Valdesolo, P., & DeSteno, D. (2006). Manipulations of emotional context shape moral judgment. *Psychological Science, 17*(6), 476–477.

Walker, M. P., & van Der Helm, E. (2009). Overnight therapy? The role of sleep in emotional brain processing. *Psychological Bulletin, 135*(5), 731–748. https://doi.org/10.1037/a0016570

Werner, G. G., Schabus, M., Blechert, J., Kolodyazhniy, V., & Wilhelm, F. H. (2015). Pre- to postsleep change in psychophysiological reactivity to emotional films: Late-night REM sleep is associated with attenuated emotional processing. *Psychophysiology, 52*(6), 813–825.

Zeelenberg, M., Nelissen, R. M., Breugelmans, S. M., & Pieters, R. (2008). On emotion specificity in decision making: Why feeling is for doing. *Judgment and Decision making, 3*(1), 18.