This document contains detailed information about the collection and processing of the data archived in the DANS-EASY archive under the title "Effects of partial sleep deprivation and threat manipulation on response inhibition, sleep and affect in healthy young adults in the region of Nijmegen, the Netherlands". The results of this study are published in 'Effects of Threat and Sleep Deprivation on Action Tendencies and Response Inhibition' (Van Peer, Gladwin, and Nieuwenhuys, *Emotion*, 2018, forthcoming).

.

Pilot Study

Prior to the main study, we conducted a pilot study among eight participants (5F/3M, age M = 24.8, SD = 3.2) to assess the feasibility and effectiveness of our sleep protocols. We used a within-subject design, in which participants followed both protocols (five hours or eight hours sleep for three consecutive nights, for details see procedure of main study below) in counterbalanced order with four normal nights in between. After the third night of each protocol, participants performed a custom-made 10-minute Psychomotor Vigilance Test (PVT; Dinges & Powell, 1985) to assess reductions in vigilance as a result of partial sleep deprivation. In this task, participants were instructed to press a button as fast as possible upon appearance of a red circle target stimulus on an otherwise black screen. Stimulus appearance was randomized with inter-stimulus intervals ranging between 2 and 10 seconds. Response time delay (in ms) was indicated by a scrolling counter and served as immediate feedback upon response. Response times below the anticipation criterion of ≤ 100 ms were excluded from further analysis. Responses without a stimulus were considered as false alarms (errors of commission), and a lapse was operationalized as a response time ≥ 500 ms.

Analyses concentrated on response times (including all single trials), with increased response times indicating decreased vigilance. Statistical analyses were performed in R with a linear mixed effects models approach (for details see Statistical Analysis of the main study

below), using a maximal random-effects structure. The model included a fixed effect for the (within-subject) factor Sleep, which was coded using a sum-to-zero contrast (5hr = 1, 8hr = -1). The repeated measures nature of the data was modelled by including per-participant random adjustments to the fixed intercept ("random intercept"), and to the slope of the within-subject factor ("random slope"). All *p*-values and confidence intervals were determined as described in the main study below (see Statistical Analyses).

Main Study

Participants

Sixty-nine students at the Radboud University Nijmegen were screened for participation in the study. Exclusion criteria were current depression (measured with the screening questions of the Major Depression Questionnaire, Van der Does, Barnhofer, & Williams, 2003) and sleep problems (indicated by a total score > 5 on the Pittsburgh Sleep Quality Index [PSQI], Buysse, Reynolds III, Monk, Berman, & Kupfer, 1989; or a total score > 2.02 on the Holland Sleep Disorders Questionnaire [HSDQ], Kerkhof et al., 2013). Five persons were excluded based on these criteria and nine (all from the five hour sleep group) dropped out prior to (n = 8) or during (n = 1) the experimental session. Of the remaining participants, three (all from the eight hour sleep group) were excluded due to insufficient adherence to the sleep protocol, as indicated by their sleep diary and Actiwatch data (see procedure below). The cut-off criterion for sleep protocol adherence was > 90 min deviation (daily average) from target sleeping time (i.e., five hours or eight hours), to avoid overlap between conditions. As a result, 52 participants were left for the analyses (n = 28 in the five hour sleep [5hr] deprivation condition and n = 24 in the eight hour sleep [8hr] control condition). All participants provided written informed consent and received course credit or financial compensation (30 euro for the full study, or 5 euro for screening and 5 euro per fulfilled day

of the sleep protocol in case of exclusion or dropout). The study was approved by the ethical committee of the Faculty of Social Sciences of Radboud University Nijmegen and was in accordance with the Declaration of Helsinki.

Procedure

Participants were recruited via the Radboud University Research Participation System (SONA) and posters and flyers in University buildings. Participants visited the lab individually at the start of the week (Monday or Tuesday) to fill in the screening questionnaires and, if they fulfilled the criteria for participation, were randomly (block method, odd-even numbers) assigned by the experimenter to the 5hr or the 8hr sleep condition. Participants in the 5hr group were instructed to sleep five hours (spend 5.5 hours in bed) in the three nights prior to the experimental session. They were advised to stay up late, rather than get up very early, and to keep bedtimes as stable as possible. Participants in the 8hr group were instructed to sleep eight hours (spend 9 hours in bed) in the three nights prior to the experimental session. Finally, participants filled in some questionnaires that measured possible confounding trait variables (Aggression Questionnaire: Buss & Perry, 1992; Dutch version Meesters, Muris, Bosma, Schouten, & Beuving, 1996; Barratt Impulsiveness Scale: Patton, Stanford, & Barratt, 1995; Dutch version Lijffijt & Barratt, 2005; Attentional Control Scale: Derryberry & Reed, 2002; Dutch version Verwoerd, Cieraad, & de Jong, 2007).

During the three-day sleep protocol, all participants engaged in normal daily activity but were asked to abstain from excessive use of alcohol and from the use of psychoactive substances that could influence their alertness or induce health risks. Participants in the 5hr group were advised not to engage in long car rides, because of the risks related to reduced reaction times and the chance of falling asleep. Furthermore, all participants were asked not to use alcohol in the 24 hours preceding the experimental session, and not to consume caffeine-containing drinks in the morning of the experimental session. Protocol adherence was checked

by means of continuous Actiwatch recording (Actiwatch 2, Philips Respironics, Murrysville, USA) and an extended version of the Consensus Sleep Diary (Carney et al., 2012), which was filled in each morning immediately after waking up (before breakfast or taking a shower).

After the third night of the sleep protocol, participants revisited the lab individually in the morning or early afternoon (i.e., on Thursday or Friday between 09:30 and 13:30) for the experimental session, in which they performed the shooting task, followed by a questionnaire to measure their subjective responses to the task.

Materials

Depression. Participants were screened for depression with the two screening questions of the Major Depression Questionnaire (MDQ, Van der Does et al., 2003). These asked whether participants experienced a period of at least two weeks in the past month in which they 1) felt almost continuously depressed or 2) lost all interest or pleasure in doing things they otherwise enjoy. Response alternatives (scored from 0-2) were 'no', 'yes but shorter than 2 weeks, namely (duration)', or 'yes'. Participants were excluded if they answered 'yes' on both questions (score = 4).

Sleep problems. A Dutch translation of the Pittsburgh Sleep Quality Index (PSQI, Buysse, Reynolds III, Monk, Berman, & Kupfer, 1989) and the Holland Sleep Disorders Questionnaire (HSDQ, Kerkhof et al., 2013) were used to screen for sleep problems. The PSQI is a self-rated questionnaire that assesses sleep quality and disturbances in the last month. It contains four questions about usual sleeping times (in CET) and ten questions about possible sleeping problems, rated on a four point scale (0 = *not at all*, 1 = *less than once a week*, 2 = *1-2 times a week*, 3 = *3 or more times a week*). From the answers seven component scores are calculated, all rated on 4-point scales (0-3): Subjective sleep quality, sleep latency, sleep duration per night, habitual sleep efficiency, sleep disturbances, use of sleeping medications, and daytime dysfunction. The total score is the sum of the seven component

scores (range 0-21). Higher scores indicate more sleep problems. The exclusion criterion for this questionnaire was a total score > 5.

The HSDQ (Kerkhof et al., 2013) is a self-assessment questionnaire for sleep disorders based on the International Classification of Sleep Disorders-2 (American Academy of Sleep Medicine, 2005). It consists of 32 questions assessing sleep disorders in the past three months, scored on a 5-point scale (1 = *not at all* to 5 = *completely*). It differentiates six symptom clusters (subscales): Insomnia, Parasomnia, Circadian Rhythm Sleep Disorder, Hypersomnia, Sleep-Related Movement Disorders, and Sleep-related Breathing Disorder. All scales were scored by calculating the average score of the relevant questions. The total score is the average score of all questions (range 1-5). The exclusion criterion for this questionnaire was a total score >2.02.

Trait variables. We measured several trait variables (aggression, impulsivity, and attentional control) that could affect inhibition, in order to control for possible confounding group differences. Aggression was measured with the Aggression Questionnaire (AQ, Buss & Perry, 1992; Dutch version: Meesters et al., 1996). This questionnaire contains 29 statements about aggression-related acts or thoughts, and respondents indicate how (un)characteristic each is for them on a 5-point scale (1 = *extremely uncharacteristic* to 5 = *extremely characteristic*). The questionnaire consists of four factors (subscales): Physical aggression, verbal aggression, anger and hostility. The total score is the sum of the factor scores (range 29-145). Incidental missing values were substituted with the average of the relevant factor (subscale) score.

Impulsivity was measured with the Barratt Impulsiveness Scale (BIS-11, Patton et al., 1995; Dutch version: Lijffijt & Barratt, 2005)). This questionnaire contains 30 questions, scored on a 4-point scale (1 = *rarely/never* to 4 = *almost always/always*). It consists of three 2nd order factors (Attentional, Motor and Nonplanning), which can be further subdivided into

six 1st order factors (Attention, Cognitive Instability, Motor, Perseverance, Self-Control and Cognitive Complexity). All factor scores are calculated as the sum score of the relevant questions. The total score is the sum of all factor scores (all questions, range 30-120).

Attentional control was measured with the Attentional Control Scale (ACS, Derryberry & Reed, 2002; Dutch version: Verwoerd et al., 2007). This questionnaire contains 20 questions, scored on a 4-point scale (1 = *almost never* to 4 = *always*). The total score is the sum of all items (range 20-80). Incidental missing items were substituted with the average score of all remaining items.

Sleep diary. To monitor sleep protocol adherence and subjective responses to the sleep protocol we used an extended version of Consensus Sleep Diary (Carney et al., 2012). This diary assessed sleep quantity and protocol adherence with six questions addressing bed times (in CET) and waking moments (number and duration). In addition, it contained nine questions asking for self-reported sleep quality, sleepiness, fatigue, fitness, feeling well-rested, alertness, positive and negative mood, and performance ability, each rated on a 10-point scale (1 = *not at all* to 10 = *very much*).

Exit questionnaire. To assess subjective responses to the shooting task and awareness of the threat contingencies, we used a custom-made self-report questionnaire, consisting of 19 questions about the shooting task (see ExitQuestionnaire_Dutch.pdf, this archive). Participants were asked which of the two opponents (represented by a picture) was associated with which sound (loud [Q4], soft [Q6]), and how certain they were about that (Q5 and Q7, respectively). Subsequently, they rated the unpleasantness of the two sounds at the beginning (loud [Q8], soft [Q10]), and end (loud [Q9], soft [Q11]) of the task, and their motivation to shoot each opponent (Q12 and Q13), on nine point Likert-scales (1 = *not at all* to 9 = *very much*). Finally, they rated their subjective responses to each opponent on nine point non-verbal pictorial scales (Self-Assessment Manikins, see Bradley & Lang, 1994. Valence (Q14,

Q17): 1 = *pleasant* to 9 = *unpleasant*; arousal (Q15, Q18): 1 = *excited* to 9 = *calm*, and dominance (Q16, Q19): 1 = *controlled* to 9 = *in control*). Additional control questions (not reported) asked for subjective feeling of control (Q1, 0 = *no control* to 100 = *total control*), estimate of the % of cases in which they felt able to prevent being shot (Q2, 0-100%), and their capacity to predict the loud noise based on the presented pictures (Q3, *yes, no, sometimes*).

Shooting Task

The shooting task consisted of an adapted (stop-signal) version of the Go/NoGo shooting task designed by Gladwin et al. (2016). It contained an introduction, training, and measurement phase. In each trial, the screen showed a view of a parking garage with an opponent character in the center of the screen, an armed police officer in the background, and a view of the participant's own "in-task" hands, holding a gun. To manipulate threat, there were two different opponents, who could be easily distinguished by their face and clothing. Both opponents behaved identically, but when participants made an incorrect response they received a loud (97 dB, 40 ms) white noise sound via headphones for one of the opponents (high threat [HT] condition, see e.g., Sperl, Panitz, Hermann, & Mueller, 2016) and a quiet sound (same sound at 50 dB) for the other opponent (low threat [LT] condition). The opponent-threat mapping was randomized across participants.

Trials began with the appearance of one of the opponents (the cue). After a variable interval (cue-stimulus interval, 0.5 to 4.5s), the opponent took one of two actions: He would draw a gun (the Go stimulus, 85% of trials) or a mobile phone (the NoGo stimulus, 15% of trials). When the opponent drew a gun, he would subsequently shoot (76% of gun trials, i.e., 65% of total trials) or he would put his gun down again (Stop signal, 24% of gun trials, i.e., 20% of total trials) after a brief delay (Stop Signal Delay, SSD). Participants were instructed to shoot the opponent, by pressing the space bar on the keyboard as fast as possible when he

drew a gun, but to inhibit this response if he put his gun down again or when he drew a phone. When participants shot the opponent in time (800 ms response window) on gun trials, they would see their own gun flash and the opponent drop down on his knees. When they responded too late, the opponent would shoot them. If participants shot before stimulus onset, after the stop signal, or in response to a phone, the police officer in the background would shoot the participant, in order to avoid strategic false-positive responding. When the opponent or the police officer shot the participants, participants would see a gun-flash and hear the loud (HT) or quiet (LT) sound. On 10% of Go trials the response window was reduced to 250 ms, so that participants would be too late and experience negative feedback (HT sound), thereby exposing them to the cue-threat contingencies even when they performed relatively quickly and accurately. Trials were separated by a variable inter-trial interval (0.6 to 0.9 s), during which the parking garage was shown without the opponents or police officer.

In the introduction phase (12 trials), participants were exposed to all possible trial types (twice with each opponent) and instructed what to do in each case. Next, participants performed a training block (100 trials, 50% LT and HT) in which all scenarios were presented in randomized order. The SSD was set at 250 ms at the start of the training, and was subsequently adjusted after each Stop signal trial as a function of participants' performance (staircase procedure, separately for LT and HT). That is, the SSD increased by 50 ms after successful stopping and decreased by 50 ms after unsuccessful stopping. The final measurement phase consisted of six blocks of 60 randomized trials each.

Data Preparation

Screening and trait questionnaires. For all screening (MDQ, PSQI, HSDQ) and trait (AQ, BIS-11, and ACS) questionnaires, total scores were calculated as described above (see references for details).

Sleep diary. Total sleep time (TST, in minutes) per night was calculated as the time between self-reported start of sleep and time of awakening, minus sleep onset latency (SOL, in minutes) and wakefulness after initial sleep onset (WASO, in minutes). In addition, the average over all nights was calculated. Incidental missing values were interpolated by averaging over remaining nights. Participants with an average TST > 90 min above or below their target sleeping time (i.e., five hours or eight hours) were excluded from the analyses. Time Awake at the moment of testing was calculated as the time (in minutes) between awakening at the morning of the experimental session (Sleep diary night 3) and the starting time of the shooting task (Time of Testing).

Actiwatch. Actigraphy data was analysed using Respironics Actiware 5 (Philips Respironics, Murrysville, USA), following the guidelines by the Society of Behavioural Sleep Medicine (see Ancoli-Israel et al., 2015). Data were visually inspected and excluded when activity counts and light values indicated detachment of the sensor. In all other cases, rest intervals were manually set when event markers (set by the participant) identified bed- and rise time, or – in case of missing event markers – when light and activity was absent. The single parameter of interest was total sleep time (TST, in minutes) per night, which was calculated as time in bed (difference between bed- and rise time) minus sleep onset latency (SOL, difference between time in bed and start of sleep) and waketime after initial sleep onset (WASO). Participants with an average TST > 90 min above or below their target sleeping time (i.e., five hours or eight hours) were excluded from the analyses.

Exit questionnaire. Objective information about which opponent was associated with the loud sound (high threat) or quiet sound (low threat) was extracted from the shooting task data. Correctness of the identification of the high and low threat opponents was calculated by comparing the opponents as identified on the questionnaire with this information. Next,

identification certainty, the motivation to shoot, valence, arousal and dominance were calculated for the high and low threat opponent (as identified based on the task data).

Shooting task. Trials with very short response windows (10% of Go trials) and trials with responses before stimulus onset (1.2% of all trials) were excluded from the behavioral analyses. Responses after the response deadline (too late, 0.9% of all trials) were coded as incorrect and were excluded from the response time analyses. Response time (RT, in ms) was calculated as the time between stimulus onset and participants' shooting responses on gun trials (correct Go response) or phone trials (false alarm response). Response accuracy was calculated as the proportion of correct responses on gun trials (correct Go responses) and phone trials (correct NoGo responses), relative to the total number of gun and phone trials, respectively. In addition, decision accuracy (sensitivity *d*') and response bias (criterion β) were calculated in SPSS following the formulas from Stanislaw and Todorov (1999), whereby "hits" were defined as correct responses on gun (Go) trials, and "false alarms" as incorrect responses on phone (NoGo) trials. Hit and false alarm rates were calculated by dividing the number of hits or false alarms by the total number of gun or phone trials, respectively. Extreme hit or false alarm rates (values of 1 or 0) were adjusted by replacing 1 with $(n - 0.5)/n$ and 0 with $0.5/n$, whereby n is the number of gun (for hit rates) or phone (false alarm rates) trials (see Stanislaw & Todorov, 1999). β values were significantly positively skewed and therefore normalized with a natural log transform before analysis (Field, Miles, & Field, 2012). Stop Signal Reaction Time (SSRT) was calculated, per participant and threat condition, with the integration method (Verbruggen, Chambers, & Logan, 2013) in Matlab. First, all RT's on gun and phone trials combined (including too late and false positive responses) were rank ordered. Then the RT value corresponding to the achieved Stop-response probability was chosen (e.g., 55^{th} percentile RT in case of unsuccessful stopping on

55% of stop trials). Finally, the SSRT was calculated by subtracting the mean SSD from this RT value. Longer SSRTs indicate decreased response inhibition.

Statistical Analyses

All statistical analyses were performed in R (R Core Team, 2016). In order to check for group differences in possible confounding variables, gender was analyzed with a Chi square test with the gmodels package (Warnes et al., 2015). Age, sum scores on sleep (HSDQ and PSQI) and trait (AQ, BIS-11, and ACS) questionnaires, Time of Testing and Time Awake at the moment of testing were analyzed with two-sided unpaired *t*-tests from the Stats package (R Core Team, 2016).

All other (repeated measures) variables were analyzed with a linear mixed effects models approach, using the glmer function (for response accuracy) or the lmer function (for all other measures) of the lme4 package (version 1.1.12; Bates, Mächler, Bolker, & Walker, 2015). We used a maximal random-effects structure (Barr, Levy, Scheepers, & Tily, 2013) for all models. The repeated measures nature of the data was modelled by including a per-participant random adjustment to the fixed intercept ("random intercept"), as well as per-participant random adjustments to the slopes of the within-subject factors ("random slopes"). In addition, we included all possible random correlation terms among the random effects.

All categorical predictors were coded using sum-to-zero contrasts. The following contrasts were used: Sleep: 5hr = 1, 8hr = -1; Threat: LT = 1, HT = -1; Stimulus: NoGo = 1, Go = -1; Night (diary and Actiwatch data): Night 1 = 1/0, Night 2 = 0/1, Night 3 = -1/-1, Time (sound ratings): Beginning of task =1, End of task= -1. All *p*-values were determined with parametric bootstrapped Likelihood Ratio Tests (X^2, using type 3 tests with 1000 simulations), performed with the mixed-function of the afex package (version 0.16.1; Singmann, Bolker, Westfall, & Aust, 2016), which in turn calls the function PBmodcomp from the package pbkrtest (version 0.4.6; Halekoh & Højsgaard, 2014). Confidence intervals were determined

using parametric bootstrapping as implemented in lme4's bootMER function, with 1000 simulations and deriving 95% confidence intervals (95% CI, type "basic") using the function boot.ci of the package boot (version 1.3.18; Canty & Ripley, 2016). Significant interactions were followed by tests of least-squares means using the lsmeans function from the package lsmeans (version 2.23.5; Lenth, 2016). Familywise error correction (*p*-value adjustment) was applied using the Tukey method, where appropriate.

To test the effects of the sleep manipulation, all sleep diary and Actiwatch data (TST and subjective responses to protocol) were analyzed (in separate models) with fixed effects for the factors Sleep (5hr, 8hr), Night (1, 2, 3), and their interaction.

To test the subjective effects of the threat manipulation in the shooting task, all ratings of the opponents were analyzed (in separate models) with Sleep (5hr, 8hr), Threat (LT, HT), and their interaction as fixed factors. The model for the unpleasantness ratings of the sounds additionally included fixed main and interaction effects for the factor Time (Beginning, End of task).

The behavioral measures of the shooting task (response accuracy, RT, and SSRT) were analyzed in separate models. All models included fixed effects for the factors Sleep (5hr, 8hr), Threat (LT, HT), and their interaction. The models for response accuracy and RT additionally included fixed main and interaction effects for the factor Stimulus (Go, NoGo). Response accuracy (weighed by the number of trials) was analyzed with a generalized model (glmer) with a binomial distribution.

To better characterize the effects on response accuracy, we performed additional exploratory analyses of signal detection measures, which allow differentiation between decision accuracy (sensitivity *d*') and response bias (criterion β). These two measures were analyzed in separate linear mixed effects models including Sleep, Threat, and their interaction as fixed effects.

Finally, as both false alarm responses on phone (NoGo) trials and SSRT are measures of response inhibition, we computed Spearman correlations (in R, with the rcorr function of the Hmisc package, Harrell Jr, Dupont, & others, 2016), separately per threat condition (LT, HT), to explore whether these two measures were correlated across participants.

As the groups differed significantly in Time Awake at the moment of testing, and this could be a confounding factor for the effects of Sleep (Thun, Bjorvatn, Flo, Harris, & Pallesen, 2015), we replicated the analyses of all main behavioral measures (response accuracy, RT, and SSRT) with Time Awake included as a control factor. We extended each of the models described above by adding a fixed effect for Time Awake (as a continuous, centered variable) and its interactions with the other predictors.

The analyses reported in the main text of the publication related to these data excluded three participants that did not adhere to the sleep protocol and – for the SSRT and correlations – two participants that did not adhere to stop instructions. The cut-off for sleep protocol adherence (>90 deviation from target sleeping time) was based on the midpoint between the target sleeping times of both groups. If a person crossed this cut-off, his or her sleeping time was closer to the target sleeping time of the other protocol than of the protocol he or she was assigned to, which invalidates the group assignment for this person. The exclusion criterion for the SSRT (stop respond rates >.20 deviation from the target rate of 0.5) was based on recommendations in the literature (see e.g., Leotti & Wager, 2010; Verbruggen & Logan, 2009) and recommendations from a colleague with ample experience with stop-signal tasks (Bram Zandbelt, personal communication). This criterion corresponded with a Z value > 2.5, which is a common cut-off value for outliers. To verify the impact of these exclusion criteria, and the robustness of our main results, we re-analyzed our behavioral data (response accuracy, RT, and SSRT) without making these exclusions. Furthermore, because including non-adherent participants causes the Sleep group factor to not reliably reflect actual sleeping

time, as explained above, we performed an additional exploratory analysis in which we included the average actual sleeping time (self-reported TST) as a standardized continuous predictor in the analyses, instead of the categorical group factor. This accounts for individual variation in adherence to the sleep protocol. The results of these analyses are reported in the Supplemental Material (Robustness Checks) of the related publication.

References

American Academy of Sleep Medicine. (2005). International classification of sleep disorders. *Diagnostic and Coding Manual*, 51–55.

Ancoli-Israel, S., Martin, J. L., Blackwell, T., Buenaver, L., Liu, L., Meltzer, L. J., … Taylor, D. J. (2015). The SBSM Guide to Actigraphy Monitoring: Clinical and Research Applications. *Behavioral Sleep Medicine*, *13 Suppl 1*, S4–S38. https://doi.org/10.1080/15402002.2015.1046356

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255–278. https://doi.org/10.1016/j.jml.2012.11.001

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Bradley, M. M., & Lang, P. J. (1994). Measuring emotion: the Self-Assessment Manikin and the Semantic Differential. *Journal of Behavior Therapy and Experimental Psychiatry*, *25*(1), 49–59.

Buss, A., & Perry, M. (1992). The Aggression Questionnaire. *Journal of Personality and Social Psychology*, *63*(3), 452–459. https://doi.org/10.1037//0022-3514.63.3.452

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

Canty, A., & Ripley, B. D. (2016). *boot: Bootstrap R (S-Plus) Functions. R package version 1.3-18.*

Carney, C. E., Buysse, D. J., Ancoli-Israel, S., Edinger, J. D., Krystal, A. D., Lichstein, K. L., & Morin, C. M. (2012). The Consensus Sleep Diary: Standardizing Prospective Sleep Self-Monitoring. *Sleep*, *35*(2), 287–302. https://doi.org/10.5665/sleep.1642

Derryberry, D., & Reed, M. A. (2002). Anxiety-related attentional biases and their regulation by attentional control. *Journal of Abnormal Psychology*, *111*(2), 225–236. https://doi.org/10.1037//0021-843X.111.2.225

Dinges, D. F., & Powell, J. W. (1985). Microcomputer analyses of performance on a portable, simple visual RT task during sustained operations. *Behavior Research Methods, Instruments, & Computers*, *17*(6), 652–655. https://doi.org/10.3758/BF03200977

Field, A. P., Miles, J., & Field, Z. (2012). *Discovering statistics using R*. London ; Thousand Oaks, Calif: Sage.

Gladwin, T. E., Hashemi, M. M., van Ast, V., & Roelof, K. (2016). Ready and waiting: Freezing as active action preparation under threat. *Neuroscience Letters*, *619*, 182–188. https://doi.org/10.1016/j.neulet.2016.03.027

Halekoh, U., & Højsgaard, S. (2014). A Kenward-Roger Approximation and Parametric Bootstrap Methods for Tests in Linear Mixed Models – The R Package pbkrtest. *Journal of Statistical Software*, *59*(9), 1–30.

Harrell Jr, F. E., Dupont, with contributions from C., & others, many. (2016). *Hmisc: Harrell Miscellaneous*. Retrieved from https://CRAN.R-project.org/package=Hmisc

Kerkhof, G. A., Geuke, M. E. H., Brouwer, A., Rijsman, R. M., Schimsheimer, R. J., & Van Kasteel, V. (2013). Holland Sleep Disorders Questionnaire: a new sleep disorders questionnaire based on the International Classification of Sleep Disorders-2. *Journal of Sleep Research*, *22*(1), 104–107. https://doi.org/10.1111/j.1365-2869.2012.01041.x

Lenth, R. V. (2016). Least-Squares Means: The R Package lsmeans. *Journal of Statistical Software*, *69*(1), 1–33. https://doi.org/10.18637/jss.v069.i01

Lijffijt, M., & Barratt, E. (2005). Dutch translation of the BIS-11. *Unpublished Manuscript, University of Houston, TX*.

Meesters, C., Muris, P., Bosma, H., Schouten, E., & Beuving, S. (1996). Psychometric evaluation of the Dutch version of the Aggression Questionnaire. *Behaviour Research and Therapy*, *34*(10), 839–843. https://doi.org/10.1016/0005-7967(96)00065-4

Patton, J. H., Stanford, M. S., & Barratt, E. S. (1995). Factor structure of the barratt impulsiveness scale. *Journal of Clinical Psychology*, *51*(6), 768–774. https://doi.org/10.1002/1097-4679(199511)51:6<768::AID-JCLP2270510607>3.0.CO;2-1

R Core Team. (2016). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.R-project.org/

Singmann, H., Bolker, B., Westfall, J., & Aust, F. (2016). *afex: Analysis of Factorial Experiments*. Retrieved from https://CRAN.R-project.org/package=afex

Sperl, M. F. J., Panitz, C., Hermann, C., & Mueller, E. M. (2016). A pragmatic comparison of noise burst and electric shock unconditioned stimuli for fear conditioning research with many trials. *Psychophysiology*, *53*(9), 1352–1365. https://doi.org/10.1111/psyp.12677

Stanislaw, H., & Todorov, N. (1999). Calculation of signal detection theory measures. *Behavior Research Methods, Instruments, & Computers*, *31*(1), 137–149. https://doi.org/10.3758/BF03207704

Thun, E., Bjorvatn, B., Flo, E., Harris, A., & Pallesen, S. (2015). Sleep, circadian rhythms, and athletic performance. *Sleep Medicine Reviews*, *23*, 1–9. https://doi.org/10.1016/j.smrv.2014.11.003

Van der Does, A., Barnhofer, T., & Williams, J. (2003). *The major depression questionnaire (MDQ)*.

Verbruggen, F., Chambers, C. D., & Logan, G. D. (2013). Fictitious Inhibitory Differences How Skewness and Slowing Distort the Estimation of Stopping Latencies. *Psychological Science*, *24*(3), 352–362. https://doi.org/10.1177/0956797612457390

Verwoerd, J., Cieraad, R., & de Jong, P. (2007). *ACS: Dutch translation of the Attentional Control Scale*. Retrieved.

Warnes, G. R., Bolker, B., Lumley, T., SAIC-Frederick, R. C. J. C. from R. C. J. are C., Program, I. F. by the I. R., NIH, of the, … NO1-CO-12400, C. for C. R. under N. C. (2015). *gmodels: Various R Programming Tools for Model Fitting*. Retrieved from https://CRAN.R-project.org/package=gmodels