

Perceiving utilitarian gradients: Heart rate variability and self-regulatory effort in the moral dilemma task

Alejandro Rosas^a, Juan Pablo Bermúdez^{b,c}, Jorge Martínez Cotrina^c, David Aguilar-Pardo^d, Juan Carlos Caicedo Mera^c and Diego Mauricio Aponte-Canencio^c

^aDepartamento de Filosofía, Facultad de Ciencias Humanas, Universidad Nacional de Colombia, Bogotá, Colombia; ^bInstitut de Philosophie, Faculté des Lettres et Sciences Humaines, Université de Neuchâtel, Neuchâtel, Switzerland; ^cCentro de Investigaciones sobre Dinámica Social (CIDS), Salud, Conocimiento Médico y Sociedad, Facultad de Ciencias Sociales y Humanas, Universidad Externado de Colombia, Bogotá, Colombia; ^dFacultad de Psicología, Universidad Católica de Colombia, Bogotá, Colombia

ABSTRACT

It is not yet clear which response behavior requires self-regulatory effort in the moral dilemma task. Previous research has proposed that utilitarian responses require cognitive control, but subsequent studies have found inconsistencies with the empirical predictions of that hypothesis. In this paper, we treat participants' sensitivity to utilitarian gradients as a measure of performance. We confronted participants (N = 82) with a set of five dilemmas evoking a gradient of mean utilitarian responses in a 4-point scale and collected data on heart rate variability and utilitarian responses. We found positive correlations between tonic and phasic HRV and sensitivity to the utilitarian gradient in the high tonic group, but not in the low tonic group. Moreover, the low tonic group misplaced a scenario with a selfish incentive at the high end of the gradient. Results suggest that performance is represented by sensitivity correlated with HRV and accompanied with a reasonable placement of individual scenarios within the gradient.

ARTICLE HISTORY

Received 28 July 2020
Revised 1 May 2021
Published online 03 June 2021

KEYWORDS

Dual-process; Heart rate variability; moral judgment; self-regulation

Moral cognition from the perspective of the neurovisceral integration model

HRV and self-regulation

Heart Rate Variability (HRV) refers to the variation in the time interval (in milliseconds) between consecutive heartbeats. Higher variability measured in a resting condition indicates a healthier state of neurovisceral integration (Porges, 1995; Thayer & Lane, 2000), indicating more robust self-regulatory capacities during emotional or cognitive challenges. According to polyvagal (Porges, 1995) and neurovisceral (McIntosh et al., 2020; Mulcahy et al., 2019; R. Smith et al., 2017; Thayer & Lane, 2000) models, bidirectional connections between cortical, limbic and brainstem structures send inputs to the sinoatrial node via the vagus nerve. These brain structures thus influence heart rate through vagal withdrawal and reactivation, producing HRV (Berntson et al., 1997). Research has established a correlation between vagal activity and changes in cerebral blood flow in the frontal lobe, which influences affective and attentional regulation (Jennings et al., 2015; Thayer et al., 2009). The associations between self-regulatory

capacity and prefrontal activity, on the one hand, and between prefrontal activity and HRV on the other, support the inference that HRV may reflect features of self-regulatory behavior. Additionally, research employing different self-regulation tasks has shown that HRV indices can serve as reliable biomarkers of self-regulation, broadly understood as the adaptive regulation of emotion, cognition and behavior in relation to environmental demands (Bridgett et al., 2015). Recent meta-analyses have supported the robustness of the link between HRV and top-down regulation in both emotional and cognitive-behavioral contexts (Holzman & Bridgett, 2017; Zahn et al., 2016).

Low indices of tonic HRV – i.e., HRV obtained from subjects in a resting state – have been associated with emotion-regulation disorders such as panic, social anxiety, and generalized anxiety disorders (Chalmers et al., 2014; Friedman, 2007), self-regulatory failures like hypervigilance (a tendency to interpret neutral stimuli as indicators of threat, Park et al., 2013), and the inability to inhibit attention to distractors while concentrating on a perceptual task (Park et al., 2014). Conversely, high indices of tonic

HRV have been linked to positive emotion and social connectedness (Kok & Fredrickson, 2010).

There is also a robust association between HRV and self-regulation in cognitive tasks. According to an influential and widely shared view, cognitive control stems from the maintenance of patterns of neural activity in prefrontal brain regions, which bias the signals of other brain structures (Miller & Cohen, 2001). Studies have found that high indices of *tonic HRV* are associated with better performance at executive-function tasks (Backs & Seljos, 1994; Capuana et al., 2014; Forte et al., 2019; Hansen et al., 2003; Luft et al., 2009; Middleton et al., 1999; Muthukrishnan et al., 2017; Park et al., 2014; Thayer et al., 2009).

The relationship between cognitive control and *phasic HRV* – i.e., HRV obtained during exposure to an emotional or cognitive challenge – is less straightforward. Studies show inconsistencies regarding whether self-regulatory effort induces phasic HRV increase or decrease relative to tonic HRV. In some tasks, self-regulatory effort is associated with a rise of phasic relative to tonic HRV. For instance, alcoholics with good impulse control over drinking underwent a clear increase in HRV during alcohol exposure, and women displayed a significant increase in HRV relative to baseline during their efforts to regulate negative emotions in a marital discussion (Ingjaldsson, Laberg and Thayer 2003; T.W. Smith et al., 2011; see also Butler et al., 2006; Park et al., 2014). In contrast, in other tasks phasic HRV suppression relative to baseline or recovery is indicative of regulatory effort (Backs & Seljos, 1994; Beauchaine, 2001; Duschek et al., 2009; Hansen et al., 2003; Middleton et al., 1999; Weber et al. c). Studies in the latter category differ with respect to whether better performance is associated with a greater or a smaller phasic suppression (Park et al., 2014). In some tasks, mostly related to working memory and attention regulation, a smaller phasic HRV suppression is associated with better performance, whereas a greater decrease in phasic HRV correlates with worse performance (Backs & Seljos, 1994; Hansen et al., 2003; Segerstrom & Nes 2007; Muthukrishnan et al., 2017). But other studies, related to sustained and selective attention or to stress regulation, have reported that greater phasic HRV suppression is associated with both better performance and higher tonic HRV (Beauchaine et al. 2007; Chapman et al. 2010; Duschek et al., 2009; Porges, 1995; Weber et al. 2010).

HRV and moral cognition

In research of moral cognition employing sacrificial moral dilemmas, participants are confronted with short stories where killing one person is conducive to saving

more lives. In such stories, a moral norm against killing conflicts with a moral norm promoting the greater good. The paradigmatic example of sacrificial dilemmas is *Footbridge*, where a large man is pushed onto the tracks of a runaway trolley to stop it from running over and killing five workers. The *moral dilemma task* consists in deciding whether it is appropriate to override the moral norm against killing for the sake of the greater good (Greene et al., 2004, 2001). Participants must assess the consequences and weigh them against the aversion toward killing. Such a task is not appropriate to test the polyvagal or the neurovisceral integration models, because it is not possible to independently identify correct performance during the moral dilemma task. In contrast to most cognitive or emotion regulation tasks, in moral dilemmas no particular response is saliently right or wrong such that it univocally indicates self-regulation. However, tests of both the polyvagal and the neurovisceral models with other self-regulation tasks confirm that high tonic HRV predicts better performance. In the moral dilemma task, this would amount to better management of the moral conflict. Accordingly, the only published HRV study with sacrificial moral dilemmas (Park et al., 2016) opted for testing whether participants better at self-regulation – with higher tonic HRV – would reveal stronger utilitarian inclinations in responses to the moral dilemmas, as claimed by the dual-process theory of moral judgment (Greene et al., 2004, 2001). According to this theory, cognitive self-regulation implies overriding a prepotent deontological emotion against killing the one person. This theory relies on research showing that utilitarian judgment in the moral dilemma task correlates with brain areas associated to cognitive control and working memory – the dorsolateral prefrontal cortex. In contrast, deontological judgment activates areas in the ventromedial prefrontal cortex and the amygdala, involved in emotional assessment (Greene et al., 2004, 2001). Park et al. (2016) found, however, that high tonic HRV predicted deontological responses, suggesting that utilitarian responders have a poorer level of neuro-visceral integration. This result contradicts the claims of the dual-process model of moral cognition. The HRV data suggest, instead, that utilitarian responders poorly integrate bottom-up emotional signals – arising in the amygdala and activating the sympathetic nervous system – with top-down neural processes. This low integration suggests some form of impairment in the processes leading to moral judgment.

The conflicting results of neuroimaging vs. HRV studies call for more research into self-regulatory processes in moral judgment. Indeed, new converging results produced by two different neuroimaging labs (Shenhav & Greene, 2014; conceptually replicated by

Hutcherson et al., 2015) suggest that good performance in moral dilemmas lies in the integration of emotional and rational inputs, not in the dominance of one over the other. These new results were obtained requesting from participants three different types of assessments in a within-subjects design: how do you feel about the options (emotional assessment); which option produces more beneficial results (utilitarian assessment); which option is more morally acceptable (integrative judgment of moral acceptability). Both labs found that the emotional and the utilitarian assessments together predicted the judgments of moral acceptability. The emotional and utilitarian assessments functioned as inputs to be pondered and integrated in an “all things considered” judgment of moral acceptability, which in both studies correlated with the activation of the ventromedial prefrontal cortex [vmPFC]. Thus, both studies suggested a new role for the vmPFC: it is no longer the locus of the emotional assessment – now ascribed to the amygdala – but is rather in charge of facilitating the *integration* of partially conflicting emotional and rational assessments. Importantly, in both studies “all things considered” judgments were spread all over a 4-point bipolar scale ranging from the appropriateness of deontological to the appropriateness of utilitarian judgment, with a mean near the midpoint. Neither study presented any claim to the effect that stronger activations in the vmPFC correlate with one particular response type (utilitarian or deontological). Absent a result of this nature, we must infer that correct performance is not tied to a utilitarian response. Good performers are, presumably, those who strive to integrate both the emotional and the “greater good” assessments, consistent with the neuro-visceral integration model of HRV (Park et al., 2016). The early efforts of Greene et al. (2001, 2004, 2008) to show that utilitarian responses are outputs of cognitive control are not aligned with the new integrative role of the vmPFC, nor with the fact that “all things considered” judgments are spread all over a 4-point bipolar scale. Conversely, these new results also cast doubt on the claim, suggested by the findings of Park et al. (2016), that good self-regulation leads to more deontological responses in the moral dilemma task.

If neither type of moral response can be excluded as expression of self-regulation, it seems appropriate to introduce a new method of measurement independent of response type. In the exploratory study reported here, we propose to measure self-regulation in relation to an objective gradient of utilitarian pull in a set of personal dilemmas. *Personal dilemmas* have two features that make the conflict salient and trigger the need for an “all things

considered” moral judgment. On the one hand, they are constructed in such a way that the agent attempting the utilitarian sacrifice must intentionally enter into physical contact with the victim (or at least perform muscular exertion) as a means to saving the others, and not merely as a side-effect of her attempt. On the other hand, if you tweak these personal scenarios in the right way, the proportion of utilitarian responses can be reliably manipulated from low to high. We explain in the overview below the meaning of these features and how they can be used to measure correct performance in participants confronting a set of dilemmas.

Overview: Sensitivity to utilitarian gradients as a measure of performance

A utilitarian gradient is realized in a set of dilemmas when the individual items depict always the same deontological cost (killing one person) and oppose it to utilitarian incentives of different magnitudes, which go from low to intermediate to high, thereby configuring a gradient in utilitarian pull. Independent studies have shown that participants, taken collectively, react rationally to utilitarian incentives, i.e., the sample mean of utilitarian response to each individual item corresponds to the item’s incentive level (Bucciarelli 2015; Christensen et al. 2014; Huebner, Hauser, & Pettit, 2011; Moore et al., 2008; Rosas and Koenigs 2014; Trémolière & Bonnefon, 2014; Gürçay & Baron, 2017). Participants are responsive to different levels of utilitarian incentives in a similar way in which, when playing a public goods game, they rationally decide their cooperative investment depending on its rate of return (Evans et al., 2015; Krajbich et al., 2015; Rosas et al., 2019). In moral dilemmas, participants estimate what can be called an “utilitarian rate of return”. We can picture this estimation as a specification of the mental process that generates an “all things considered” judgment of moral acceptability. In this sense, measuring sensitivity to utilitarian gradients would allow us to see integrative moral judgment in action.

We thus thought it reasonable to explore whether HRV relates in the expected way to participants’ sensitivity to the gradient. A reasonable hypothesis is that differences in tonic (resting) HRV would relate to differences in gradient sensitivity. Participants with higher tonic HRV would exhibit, e.g., greater sensitivity; or alternatively, they would perceive the ordering of scenarios in the gradient in closer conformity to the ordering of costs and benefits. Similarly, we would expect that HRV would either decrease or increase during the task (phasic HRV). This would replicate what has been observed in

many experiments confronting participants with emotional or cognitive challenges. Since dilemmas contain both emotional and cognitive challenges, we did not make a more specific prediction, in accordance with the exploratory nature of this study.

It is worth emphasizing that this method is not subject to the arbitrariness affecting the claim that either utilitarian or deontological responses index correct performance. It is not unreasonable to expect that participants have different thresholds at which for them the balance tips in favor of the utilitarian option. But if some participants lack such thresholds or show inconsistencies in the placement of scenarios along the gradient, it would be interesting to see whether this behavior somehow relates to tonic and/or phasic HRV levels. Attending to the latter possibility, we included in the set of scenarios one that involves a risk to the agent herself: in this scenario (labeled *Flames*, see [Appendix](#)), the agent (who is always identified with the participant) must sacrifice one individual to save five others, and the agent is one of them. Since low HRV is correlated with anxiety, panic, and a heightened sense of threat, high- and low-HRV participants might well respond differently to a scenario where their own (hypothetical) survival is at stake.

In this sense, sensitivity to the gradient indicates correct performance in a way that is free from the arbitrariness of considering either type of judgment to be generally correct. The method uses a purely formal feature of the cognitive process underlying the formation of “all things considered” moral judgments, namely, sensitivity to a utilitarian gradient. Note also that, although we talk of an “utilitarian gradient”, the gradient is also deontological in the opposite direction, as is natural in a bipolar scale.

Methods

Participants

We recruited 82 advanced undergraduate and graduate participants by word of mouth from universities in Bogotá, Colombia. Participants were 21–26 years of age, $M_{age} = 24.8$, $SD = 1.45$. This strict age limit was established to avoid known developmental confounds on HRV (Sosnowski, 2010). Proportion of females was 54%. The study was conducted in Spanish (see [Appendix](#) for original materials and translations). Participants refrained from alcohol and drug use for 4 h prior to their participation. They were also screened for a history of neurological disease. None reported any antecedents. Eleven participants had incomplete measurements due to errors in the digital recording of their electrophysiological data

and two participants had incomplete utilitarian responses. This yielded a final $n = 71$ participants for the ECG measures and 80 participants for the utilitarian responses.

Procedure

All participants were tested individually in a well-lit room at the Interdisciplinary Laboratory of Human Sciences and Processes at the Universidad Externado de Colombia. Participants were welcomed and asked to read and sign the informed consent form, and to fill a survey that included health and socioeconomic questions, and a question about religiosity. Participants were then fitted with the electrocardiogram (ECG) electrodes. After checking for proper equipment functioning and signal registration, participants were asked to remain in a resting position for 5 minutes, during which baseline measurements were taken.

When the baseline period was completed, we presented participants with the moral dilemma task. SuperLab was used to present the tasks and measure response times, while synchronizing these response times with electrophysiological signals. We asked participants to judge each situation presented to them from a moral point of view, imagining that the actions entailed no harm to themselves and had no legal or other consequences. Participants judged five scenarios in a within-subjects design. ECG measurements were taken for each of the five scenarios, which were presented sequentially and counterbalanced. For each scenario, the ECG measurements were taken in three self-paced phases: 1. the initial dilemma description, 2. the presentation of the utilitarian option, and 3. the request for a response. Participants moved through the three phases of each dilemma by pressing the space bar; this action was synchronized with ECG measurements, marking the beginning and the end of each phase. Phasic HRV was measured in the *response* phase, which spanned from the presentation of the question to the participant’s registration of an answer. To illustrate, we present here the three phases in the scenario *Shark*, which in our study represents the baseline dilemma with the lowest utilitarian pull:

Description: You are watching an exhibition of sharks being fed in an aquarium pool. A metal fence suddenly collapses, and a group of people fall into the pool. Their frantic movement in the water attracts the hungry sharks. A person next to you has a harmless episode of nosebleeding.

Options: If you push this person into the water, their blood will draw the sharks away from the other five people. The bleeding person will die, but the five people will swim to safety.

Response: How right or wrong is it to cause the death of the person in order to save five people?

Participants registered their responses in a 4-point scale (1 = “Entirely wrong”, 2 = “More wrong than right”, 3 = “More right than wrong”, 4 = “Entirely right”). After participants responded to all five dilemmas, electro-physiological measurements stopped. All statistical analyses were carried out with SPSS version 25.0. The data are available at <https://osf.io/vdrzm/>

Materials

Sacrificial dilemmas

We designed a sequence of five dilemmas using only high-conflict personal scenarios, i.e., scenarios where the sacrifice of one person involves contact with the victim and/or exertion of muscular force and is knowingly performed as a means, not merely as a side-effect, to saving others.

Our scenarios included features that would confront participants with different utilitarian rates of return. *Shark* was our *Footbridge*-like scenario and acted as the baseline. It was constructed with four features: (1) a modest 1:5 kill-save ratio; (2) the victim is innocent of the threat of imminent death impending over others; (3) the victim will not die unless sacrificed to save the five; and (4) the agent performing the sacrifice has no personal stake in the sacrifice. In each of the other four scenarios we tweaked a single one of these four features to give the scenario a stronger utilitarian pull: *Dam* changed the kill-save ratio, i.e., one hundred thousand people would be saved by sacrificing one innocent person; in *Grenade* the person to be sacrificed is not innocent, because he credibly threatens five innocent others with imminent death; *Boat* features an innocent victim who would die along with five other persons even if no action was taken, but whose sacrifice would save the other five; and *Flames* featured an agent that would save herself along with four others by sacrificing one innocent person. We included these differences to create a utilitarian gradient, taking into account previous research that shows their effects in the percentages or means of utilitarian response (Christensen et al. 2014; Trémolière & Bonnefon, 2014; Rosas & Koenigs 2014; Bucciarelli, 2015; Rosas et al., 2019). The results were as expected: across the five scenarios we obtained three significantly distinct levels of mean utilitarian responses in the four-point scale.

Physiological measurements

We used PowerLab 16/35 ADC to record electrophysiological measures, amplified with a DualBio amplifier.

Recordings were sampled by PowerLab at 2000 Hz. Superlab installed on a Dell personal computer (Intel Core i7-6700, 3.40 GHz, 16 GB RAM) was used to show stimuli to participants. Superlab sent stimuli markers via USB to Cedrus StimTracker, which passed them to Powerlab via digital port. PowerLab was connected via USB to a second Dell personal computer (Intel Core i7-6700, 3.40 GHz, 16 GB RAM) running LabChart. These devices and connections ensured fidelity and Jitter-free markers in the recordings.

Using Labchart’s built-in software filters, we filtered ECG/EKG data applying a digital Notch Filter to remove power line noise in the 60 Hz frequency band. We also used a high pass filter of 0.5 Hz with a 2nd-order Butterworth, internal parameters cutoff frequency of 0.26 and a transition width of 1.21. Finally, we applied a low-frequency filter of 20 Hz with a 4th-order Butterworth, internal parameters cutoff frequency of 22.92 and a transition width of 25.6. To obtain HRV measures, Labchart provided automatic beat recognition based on R wave detection and a powerful classifier for normal and ectopic beats. The R waves detected were visually inspected by two researchers. We also used Labchart to calculate HRV measures in the time and frequency domains from the intervals of the markers based on the detected R waves.

ECG provides inter-beat interval (IBI) information that specifies the temporal distances between each R-spike and the next (R-R intervals). To derive HRV from this information, researchers can apply time-domain or frequency-domain approaches. Time-domain approaches apply statistical computations to the R-R intervals to derive information about IBI variance. Among the most common time-domain measures are the root mean square of successive differences of successive intervals (*RMSSD*) and the percentage of adjacent normal-to-normal intervals that differ from each other by more than 50 ms (*pNN50*). Frequency-domain approaches split ECG data into different frequency components of the power spectral density band to provide HRV estimates within specific bands. The most commonly used frequency-domain measures are the high-frequency component (*HF*, between 0.15 and 0.4 Hz), the low-frequency component (*LF*, between 0.04 and 0.15 Hz) and the *LF/HF ratio*. (Task Force of The European Society of Cardiology and The North American Society of Pacing and Electrophysiology, 1996; Thayer et al., 2010).

Among the frequency-domain components, there is general agreement that HF represents vagally-mediated parasympathetic nervous system activity. Thus, we focus our frequency-domain analyses on HF measures. The parasympathetic system’s activity registers prefrontal activity

through the vagus nerve and allows us to theoretically connect HRV to self-regulatory capacity (Hansen et al., 2003; Holzman & Bridgett, 2017). In measuring HRV in the HF band (RSA), we did not control for respiration frequency, as some researchers recommend (Grossman et al., 1991). According to a review of the literature (Denver et al., 2007), there is no evidence of a lawful connection between RSA amplitude and frequency of respiration. Denver et al. (2007) also confirmed with two experiments that this holds true both in a resting condition and under manipulation of the amplitude of RSA via atropine.

Results

The utilitarian gradient in the five scenarios

Our five scenarios configured a gradient of utilitarian response. A related samples Friedman's analysis of variance by ranks with Bonferroni post hoc comparisons confirmed three significantly distinct levels of mean utilitarian response (all $ps \leq 0.01$). *Shark* and *Grenade*, the low and high end of the gradient, respectively, were significantly different from each other and from *Boat*, *Dam* and *Flames*. The latter three items occupied the middle level and were not significantly different from each other but differed significantly from both the low and high ends of the gradient (Figure 1 and Table 1).

Utilitarian response correlates negatively with the HRV measures

The HRV measures of HF, RMSSD and pNN50, for both the tonic and the phasic HRV, were log10 transformed to obtain a normal distribution. We first checked for age

effects on tonic HRV with Pearson correlations (2-tailed) and found none (all $ps > .09$). We computed the total utilitarian response (bipolar scale) to the five dilemmas for each participant. This variable originated from the ordinal 4-point scale of utilitarian response and was treated as ordinal. We then ran Spearman rho correlations (2-tailed) between total utilitarian response and HRV measures. We successfully replicated the key finding in Park et al. (2016): utilitarian response was negatively correlated with the three tonic HRV measures (HF, RMSSD and pNN50). The effect sizes of the correlations with the tonic measures were roughly of the same magnitude as those found by Park et al.: $-.227 \leq rs \leq -.343$, $p < .001$. We also found negative correlations between utilitarian response and the three phasic HRV measures (marginal with phasic HF, $p = .052$, see table S1 in the Appendix).

Effects of tonic and phasic HRV on moral sensitivity

We calculated the response slope of the regression line for each participant with the slope function in Excel, ordering scenarios according to their sample means as shown in Table 1. The average regression slope for the sample was .25, $SD = .19$. Roughly, the average slope represents the sample's average moral sensitivity to the utilitarian gradient, as defined above. Since we have conceded that differences in moral sensitivity are legitimate, it is not possible to establish a unique correct or optimal slope. But plausibly, everything far above and far below one standard deviation from the average could count as incorrect performance; moreover, large divergences in the placement of one or several scenarios should be flagged for interpretation of performance, in

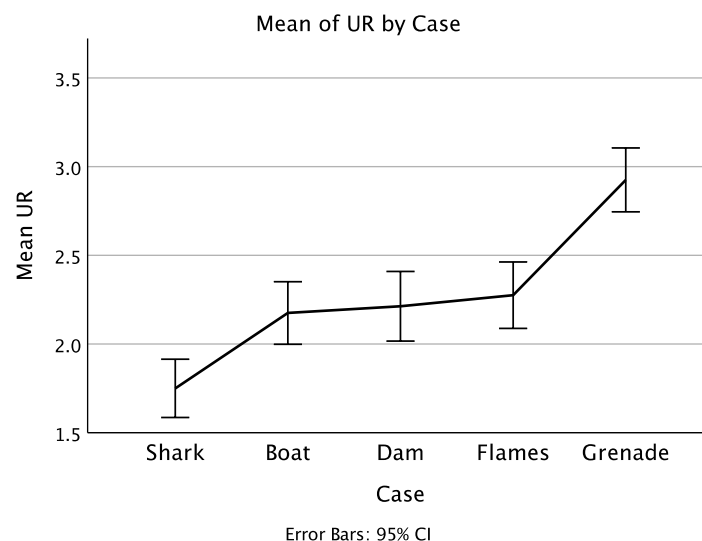


Figure 1. The five scenarios build a gradient of utilitarian response in a 4-point response scale.

Table 1. Descriptive statistics: Mean and SD of utilitarian response.

Case	<i>N</i>	Mean	Std. Deviation
Shark	79	1.73	.729
Boat	80	2.18	.792
Dam	80	2.21	.882
Flames	80	2.28	.842
Grenade	80	2.93	.808

conjunction with HRV data. Participants' slopes ranged between $-.3$ and $.8$, including 0 . Slope $-.2$ was not represented in any participant and some of the slopes were sparsely populated (see histogram, Figure 2a). However, we did not exclude any slope data from the following analyses. Since the participants' slope was derived from the ordinal utilitarian response in a 4-point scale, we treated the variable as ordinal in the correlation analyses. Spearman rho correlations (two-tailed) between moral sensitivity and the three measures of log10 transformed HRV, both tonic and phasic, revealed no correlation of sensitivity with tonic, and a small positive correlation of sensitivity with phasic HRV on the three measures. The correlation values with phasic HF, RMSSD and pNN50 were, respectively: $r_s = .153$, $p = .017$; $r_s = .139$, $p = .008$; $r_s = .182$, $p = .011$. In the correlation with HF, to ensure the detection of 0.15 Hz in the high-frequency spectrum, we excluded the data points where participants responded in less than 6.7s to any one of the 5 dilemmas (105 of 356 valid data points). We also ran the analysis including those data points and the result was similar: no correlation with tonic and a small positive correlation with phasic HF ($r_s = .109$, $p = .04$).

As mentioned above, in some cognitive task studies neither performance nor HRV-change relative to baseline are uniform across high vs. low tonic HRV (Weber et al. 2010; Park et al., 2014). Given this, we explored whether the links between performance (as measured by moral sensitivity) and HRV-change could be different for participants with high vs. low tonic HRV. Thus, we repeated the correlation analyses separately for high vs. low tonic groups. The two groups were based on the median split of log10-transformed tonic HRV measures. In the high tonic group, we obtained medium-sized positive correlations between moral sensitivity and both tonic and phasic HRV. The r_s values for the high tonic group were, with tonic: (HF) $r_s = .493$, $p < .001$; (RMSSD) $r_s = .256$, $p = .001$; (pNN50) $r_s = .298$, $p < .001$; and with phasic: (HF) $r_s = .379$, $p < .001$; (RMSSD) $r_s = .306$, $p < .001$; (pNN50) $r_s = .249$, $p = .003$. In the low tonic group, we observed tendencies toward a negative correlation between tonic HRV and moral sensitivity. In that group, the positive correlations

between sensitivity and phasic HRV were not statistically significant, except for a small correlation with HF: $r_s = .194$, $p = .037$.

Although the median split is an arbitrary method to separate low vs. high tonic groups, it reveals that the correlation between HRV and moral sensitivity is not uniform across all levels of tonic HRV. It is almost inexistent in the low tonic group, but it is robust at higher levels. A plausible interpretation – in line with previous research on the relation between self-regulation and HRV – points to more controlled moral sensitivity at higher tonic levels. At these levels, individuals exert effort to carefully estimate the utilitarian rate of return in the moral dilemma task, while at low tonic levels responsiveness to utilitarian gradients is likely based on intuitions prone to distortions. We comment on this in the discussion. A significant change from tonic to phasic HRV would indicate this difference in effort. We focus the remaining analyses on the log10 transformed high frequency (HF) variable.

We plotted the mean tonic and phasic HF values on separate lines against the slope scores on the x-axis (Figure 2b), separating the plots in low ($N = 122$ responses) vs. high ($N = 129$ responses) tonic HF groups. Figure 2a shows that the slope range between 0 and $.4$ is the most populated and thus yields the more reliable data. In the segment from $.2$ to $.4$, Figure 2b visually displays the positive correlation between slope and phasic and tonic HF in the high tonic group, and the absence of such correlation in the low tonic group. In the high tonic group, a statistically significant suppression of phasic HF accompanies its positive correlation with slope. We confirmed this with a comparison of mean tonic vs. phasic HF for the high tonic group in the slope range ($0, .4$). Tonic HF ($M = 2.75$, $SD = .49$) was significantly higher than phasic HF ($M = 2.56$, $SD = .69$), $t(224) = 5.08$, $p < .001$.

We then ran linear regressions with slope as DV. In the high tonic group, phasic HF was a significant predictor of slope when entered as the sole IV in the regression: $Beta = .244$, $t(123) = -2.79$, $p = .006$, $adj. R^2 = .05$. When entered together with tonic HF, only the latter was a significant predictor: $Beta = .502$, $t(122) = 5.44$, $p < .001$, $adj. R^2 = .23$, indicating that tonic HRV accounted for 23% of the variance in moral sensitivity; the values for phasic HF were: $Beta = -.017$, $t(122) = -.19$, $p = .85$. The effect of phasic HF on slope in the high tonic HF group was thus completely mediated by tonic HF. In contrast, neither tonic nor phasic HF predicted slope in the low tonic group, neither when entered alone nor together (all $ps > .23$).

Despite the important differences between the low vs. high tonic groups in the relationships between tonic HF,

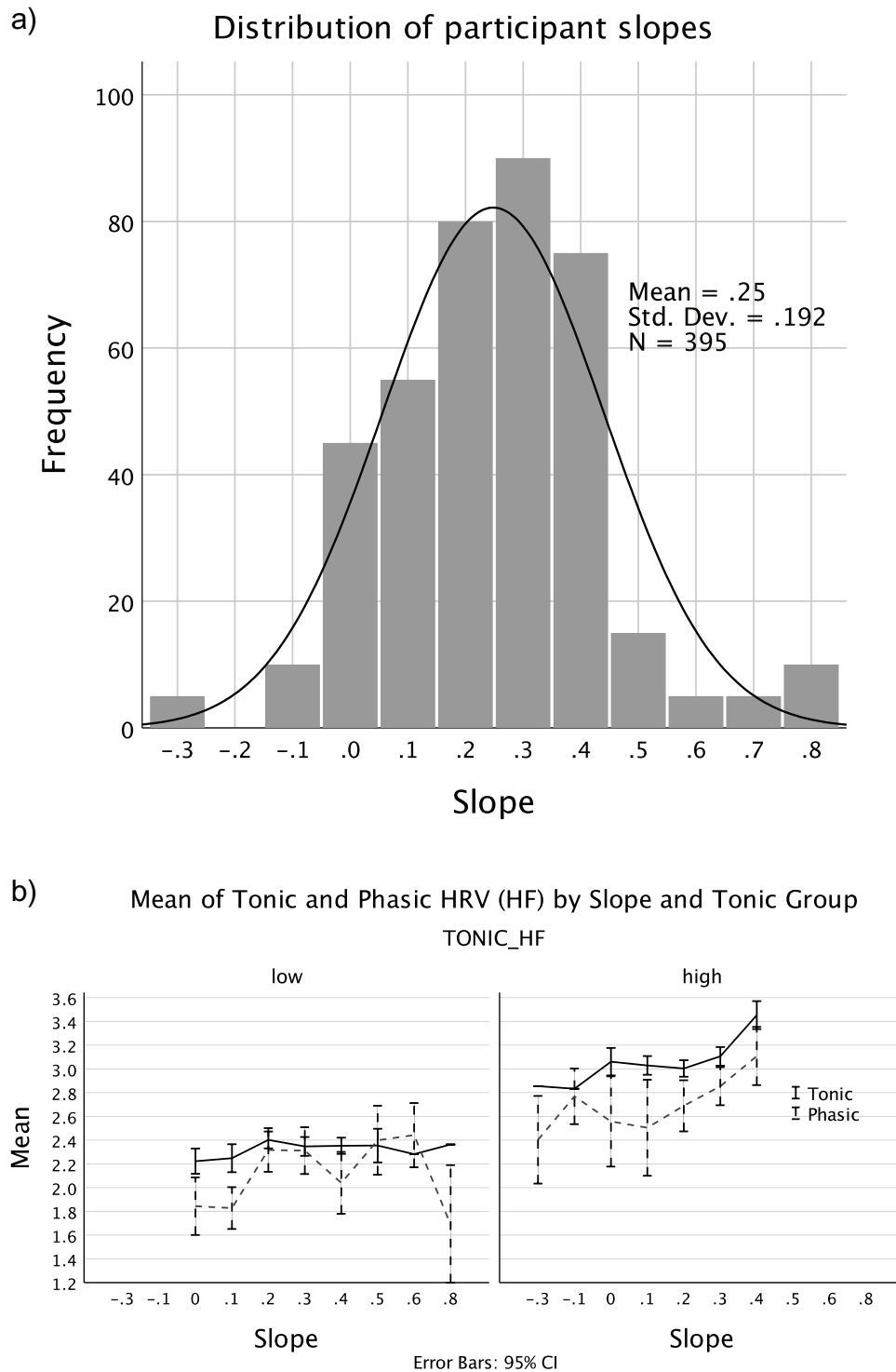


Figure 2. a) Distribution of participant slopes. The frequency represents responses to each individual dilemma, where each participant delivered 5 responses. The slope -0.2 was not represented in any participant. The slope 0.7 was represented in a participant with no HF data. b) Mean tonic and phasic HF as a function of participant slope and high ($N = 129$ responses) vs. low ($N = 122$ responses) tonic group. Groups were based on the median split of \log_{10} transformed tonic HF, after subtracting 105 data points from responses where participants took <6.7 sec.

phasic HF, and slope, the low tonic group also exhibited moral sensitivity corresponding to the slopes between 0 and .4. We thus explored differences between the low vs. high tonic HF groups regarding their placement of

scenarios along the utilitarian gradient. To compare the two groups in this respect, we plotted the mean UR for each scenario for each group. The groups differed in only one respect: their attitude to *Flames*, the only scenario in

our battery where the self-preservation of the participants was hypothetically at stake. The high tonic group rated it as of low utilitarian pull (second to lowest after *Shark*) while the low tonic group rated it as second to highest after *Grenade* (Figure 3). Independent samples t-tests reveal a difference in mean UR in low vs. high tonic group only in *Flames*: $t(51) = 2.570$, $p = .013$. Consistently, the negative correlation found in the whole sample between tonic HF and UR ($r_s = -.227$, $p < .001$) disappeared when we repeated the analysis excluding *Flames* ($r_s = -.088$, $p = .219$). We comment on this finding in the discussion.

No significant relation of total UR or reaction time to the high vs. low tonic group

Finally, we explored whether total UR and reaction time related to tonic and phasic HF in a similar way as slope does, i.e., whether these variables exhibited particularly significant relations to the high vs. the low tonic groups. Reaction time did not correlate significantly with either tonic or phasic HF. When we divided the sample in low vs. high tonic groups based on the median split of log10 transformed HF, no significant results emerged for either total UR or reaction time. Total UR's negative correlations with both tonic and phasic HF emerged only for the whole sample ($r_s = -.227$, $p < .001$, and $r_s = -.125$, $p = .052$, respectively), and disappeared when looking at either group of the divided sample.

Discussion

Research into moral judgment using HRV measures and the moral dilemma task is hampered by the lack of an objective measure of performance in that task. The only study to date addressing this subject found that utilitarian judgment was negatively correlated to tonic HRV, suggesting that utilitarian responders have a poorer level of neuro-visceral integration (Park et al., 2016). This result contradicts the claims of the original dual process model of moral cognition (Greene et al., 2008, 2004, 2001), namely, that utilitarian responses express the ability to exert cognitive control over a prepotent emotional response. We do not deny that there is a reasonable level of certainty in the inference from levels of tonic HRV to performance on any cognitive task, and thus also in the moral dilemma task. However, with respect to the moral dilemma task, an independent measure of performance would provide research with a more fine-grained picture of regulatory effort in that task. Discriminating performance on the basis of moral response type is not an option, given that neither deontological nor utilitarian responses can be straightforwardly classified as correct or incorrect. Moreover, new neuroscientific results (Hutcherson et al., 2015; Shenhav & Greene, 2014) suggest that the vmPFC integrates emotional and rational assessments into "all things considered" moral judgments. Since these *integrative* moral judgments are spread evenly across a 4-point bipolar scale, it seems out of the question to identify correct performance, i.e., successful integration, with any of the two moral response types.

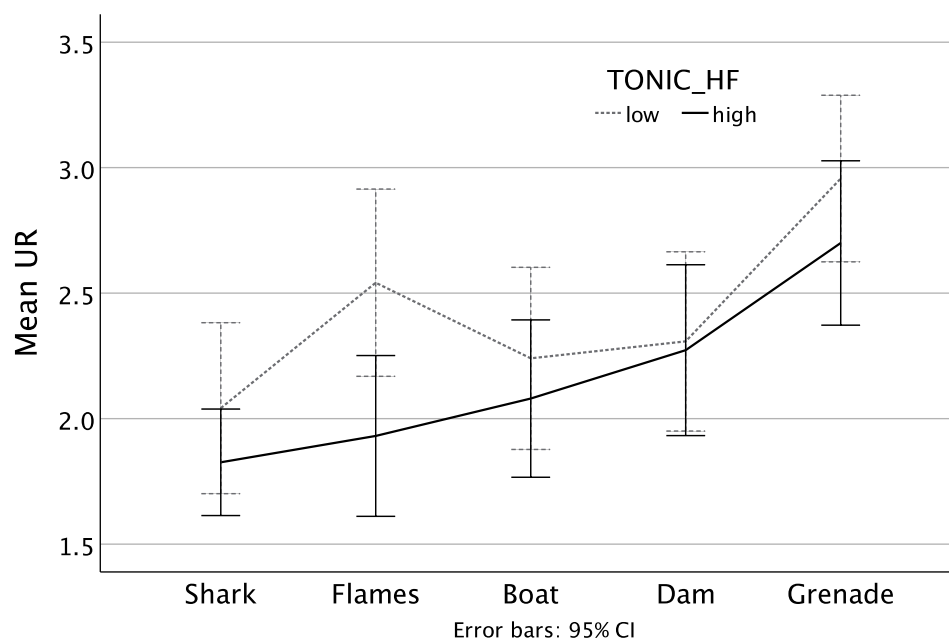


Figure 3. Perception of the utilitarian gradient as a function of level of tonic HF.

In this paper, we provide an independent measure of self-regulation for this task. To achieve this, we link the new neuroscientific evidence about integrative moral judgments (integrating emotional “harm aversion” and rational “greater good” estimates) with a feature of personal moral dilemmas that has been replicated across many studies. Although the paradigmatic personal dilemma – namely *Footbridge* – usually receives a low percentage of utilitarian approval, its characteristic features can be tweaked to create statistically significant increments in utilitarian responses. Those features change the utilitarian “rate of return”, i.e., the participants’ estimate of how justifiable it would be to violate a norm against killing in order to save lives (Moore et al., 2008; Christensen et al. 2014). It is reasonable to expect from participants some moral sensitivity to this “rate of return”. Thus, we confronted them with a set of five moral dilemmas that exhibit a gradient of low to high rates. The gradient is initially set by ordering the sample’s mean responses to the different scenarios from low to high. We represent the participants’ sensitivity to the utilitarian gradient as their response slope to the five scenarios, and we use the participants’ sensitivity as a measure of self-regulation.

In tasks where performance is positively correlated with regulatory capacities, higher tonic HRV is expected to correlate with better performance. Thus, in this study, sensitivity to the utilitarian gradient would be expected to correlate with higher tonic HRV. Non-parametric correlations showed no relation between tonic HRV and participants’ response slope to the set of dilemmas. But after dividing the sample into high vs. low tonic HRV groups, we observed medium-sized positive correlations between three measures of tonic HRV and participants’ response slope *only* for the high tonic group. Apparently, only high tonic HRV participants make a controlled estimate of the utilitarian rate of return. Thus, only their response slope fits our measure of performance.

Another expectation is that differences in regulatory capacity should make a difference with respect to phasic HRV reactivity measures. Regulatory effort was expressed as phasic HRV suppression in the high tonic group. A statistically significant decrease in phasic HF was observed in the slope range between 0 and .4. In the low tonic group, participants show no consistent pattern in phasic reactivity. Moreover, although we had observed small positive correlations between slope and phasic HRV measures in the whole sample, they were medium-sized in the high tonic group, while they were absent in the low tonic group. This suggests that the ability to discriminate differences in the utilitarian pull increased with the regulatory effort exerted by participants in the high tonic HRV group, and that only high-

tonic participants exerted the effort. Any responsiveness shown by low tonic HRV participants was probably intuitively guided and subject to distortions. These results present evidence for considering sensitivity to a utilitarian gradient across personal dilemmas as a measure of performance.

There are two issues that could cause concern. One is minor, namely, how to interpret the negative slopes (–.3 and –.1) observed in two high tonic participants. Adequate sensitivity to the utilitarian gradient should preclude negative slopes. We believe it is unlikely that this observation carries any significance. But larger samples and comprehension checks would be needed to decide this issue.

The second concern is critical. Although the pattern of relations between slope, tonic HF and phasic HF is consistent with HRV findings in other cognitive tasks, this pattern emerged only in the high tonic group. The low tonic group also populated the slope range that indexes correct performance without exhibiting the expected pattern. This suggests that the slope range alone cannot pin down regulatory failure. It is thus fitting that we did observe a difference between the two groups regarding the scenario ordering. These groups placed *Flames* on opposite sides of the gradient (Figure 3). This is no small difference, since *Flames* differs from the baseline scenario only by including a selfish stake in the utilitarian action. Low tonic participants who rate this scenario at the high end of the gradient are susceptible to a distortion in their moral judgment, perhaps due to greater levels of personal anxiety. Therefore, regulatory failures cannot be spotted using slope values alone. Instead, they are primarily identified in failures to discriminate between utilitarian moral and other motivations (e.g., self-preservation).

This last point has interesting implications regarding the focus of future research with moral dilemmas. Research has been all too narrowly focused on the opposition between deontology and utilitarianism and has created the expectation that performance is somehow linked to it. But plausibly, low vagal tone would not produce an increase in utilitarian responses unless some scenario or scenarios were overrated, and thus misplaced along the gradient. What previous research interpreted as a poor neurovisceral integration in utilitarian inclinations expressed in moral dilemmas, might be more properly viewed as an excessive influence of non-moral incentives on moral judgment, caused by self-regulatory failures in participants with low vagal tone. This kind of confusion distorts the integration of conflicting moral considerations, which was postulated as the role for the vmPFC

by the new neuroimaging studies reviewed in the introduction.

This result points us to a plausible conclusion: low vagal tone causes failures in the estimation of the utilitarian rate of return of some scenarios by allowing non-moral features to influence the moral integration process, rather than by simply failing to instantiate moral sensitivity. Further research should try to establish how these failures arise. One possibility, compatible with our data, is that participants initially produce an intuitive estimation of the rate of return, i.e., the “all things considered” moral judgment. This intuitive processing might yield roughly correct estimates in many cases. But some scenarios might pose a need to cross-check the results via effortful deliberation. This second process might fail to intervene when vagal tone is low and would be responsible for the misplacement of some scenarios.

Limitations and recommendations for future research

In this study, we attempted to avoid well-known developmental confounds on HRV by including only young adult participants (age range = 21 to 26 y.). We did not, however, measure other possible factors like exercise level or socioeconomic status, which could possibly affect HRV levels. Those factors function as causes of high or low levels of tonic HRV. We left out the causes to focus on whether HRV can be used as a measure of moral performance. Since utilitarian responses do not seem to index self-regulatory effort – utilitarian scores are negatively correlated with tonic HRV levels – we explored whether sensitivity to a utilitarian gradient could function as such a measure. Our study shows that participants’ slopes in a range from 0 to .4 and a morally acceptable placement of items within the utilitarian gradient correlates with high tonic HVR in a moral dilemma task.

This measure does not point to utilitarian responses as uniquely indexing self-regulatory effort, which was proposed by the original dual-process theory of moral cognition (Greene et al., 2004, 2001). If our performance measurement proposal is correct, future research will need to re-conceive the contributions of intuition (system 1) and deliberation (system 2) (Evans & Stanovich, 2013) to moral judgment. Moral dilemmas might, for example, activate intuitions of both utilitarian and deontological principles. Some recent research points in this direction (Bago & De Neys 2019; Rosas & Aguilar-Pardo, 2020). After moral intuitions are activated, participants would engage deliberation to confirm whether the

intuitive answer properly accommodates all the features of the situation. However, participants can and will differ in the response they reach after deliberation. This is a plausible hypothesis that could be used to guide future research.

The use of scenario-placement along a utilitarian gradient as a measure of performance introduces some additional complication in the methods, for researchers will need to agree as to what are unacceptable slopes and unacceptable item-placements. That said, in some cases, those agreements are not difficult to reach, as is the case of a scenario that is baseline in all aspects, except that the agent has a selfish stake in the sacrifice. And once agreements are reached, it is easy to measure performance. The measure proposed here is compatible with possible instances of *deontological* responses to a selfish stake scenario being the result of self-regulatory effort. Though it is reasonable to be flexible regarding which utilitarian incentives override a deontological norm against killing, some motivations – e.g., a self-preservation anxiety – could not be meaningfully taken to be overriding moral factors.

An important limitation in our study is that we did not implement an independent control task of self-regulation. This could be done with a self-report questionnaire, e.g., the Difficulties in Emotion Regulation Scale (DERS) (Gratz & Roemer, 2004) or the Self-Control Scale (Tangney et al., 2004), or perhaps better, by requesting participants to complete a separate cognitive task with correct and incorrect answers, which would allow researchers to measure HRV and self-regulation independently of the moral dilemma task and to compare it with the moral sensitivity to utilitarian gradients. This control task should be implemented in future research.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This project was funded by the Research Division of the Universidad Nacional de Colombia, project 37159, 2017–2018, and by the Universidad Externado de Colombia, project 370072017–2018. It was supported by the Swiss National Science Foundation’s project The Nature and Value of Efforts.

ORCID

Juan Pablo Bermúdez  <http://orcid.org/0000-0001-5239-2980>

References

- Backs, R. W., & Seljos, K. A. (1994). Metabolic and cardiorespiratory measures of mental effort: The effects of level of difficulty in a working memory task. *International Journal of Psychophysiology*, *16*(1), 57–68. [https://doi.org/10.1016/0167-8760\(94\)90042-6](https://doi.org/10.1016/0167-8760(94)90042-6)
- Beauchaine, T. (2001). Vagal tone, development, and Gray's motivational theory: Toward an integrated model of autonomic nervous system functioning in psychopathology. *Development and Psychopathology*, *13*(2), 183–214. <https://doi.org/10.1017/S0954579401002012>
- Beauchaine, T., Gatzke-Kopp, L., Mead, H., 2007. Polyvagal theory and developmental psychopathology: emotion dysregulation and conduct problems from preschool to adolescence. *Biological Psychology*, *74*(2), 174–184.
- Berntson, G., Bigger, J., Eckberg, D., Grossman, P., Kaufmann, P., Malik, M., Nagaraja, H., Porges, S., Saul, J., Stone, P., & Van Der Molen, M. (1997). Heart rate variability: Origins, methods, and interpretive caveats. *Psychophysiology*, *34*(6), 623–648. <https://doi.org/10.1111/j.1469-8986.1997.tb02140.x>
- Bridgett, D. J., Burt, N. M., Edwards, E. S., & Deater-Deckard, K. (2015). Intergenerational transmission of self-regulation: A multidisciplinary review and integrative conceptual framework. *Psychological Bulletin*, *141*(3), 602–654. <https://doi.org/10.1037/a0038662>
- Bucciarelli, M. (2015). Moral dilemmas in females: Children are more utilitarian than adults. *Frontiers in Psychology*, *6*, 1345. <http://dx.doi.org/10.3389/fpsyg.2015.01345>
- Butler, E. A., Wilhelm, F. H., & Gross, J. J. (2006). Respiratory sinus arrhythmia, emotion, and emotion regulation during social interaction. *Psychophysiology*, *43*(6), 612–622. <https://doi.org/10.1111/j.1469-8986.2006.00467.x>
- Capuana, L. J., Dywan, J., Tays, W. J., Elmers, J. L., Witherspoon, R., & Segalowitz, S. J. (2014). Factors influencing the role of cardiac autonomic regulation in the service of cognitive control. *Biological Psychology*, *102*(1), 88–97. <https://doi.org/10.1016/j.biopsycho.2014.07.015>
- Chalmers, J. A., Quintana, D. S., Abbott, M. J. A., & Kemp, A. H. (2014). Anxiety disorders are associated with reduced heart rate variability: A meta-analysis. *Frontiers in Psychiatry*, *5* (JUL), 1–11. <https://doi.org/10.3389/fpsyg.2014.00080>
- Chapman, H. A., Woltering, S., Lamm, C., & Lewis, M. D. (2010). Hearts and minds: Coordination of neurocognitive and cardiovascular regulation in children and adolescents. *Biological Psychology*, *84*(2), 296–303. <https://doi.org/10.1016/j.biopsycho.2010.03.001>
- Christensen, J. F., Flexas, A., Calabrese, M., Gut, N. K., & Gomila, A. (2014). Moral judgment reloaded: A moral dilemma validation study. *Frontiers in Psychology*, *5*(JUL), 1–18. <https://doi.org/10.3389/fpsyg.2014.00607>
- Denver, J. W., Reed, S. F., & Porges, S. W. (2007). Methodological issues in the quantification of respiratory sinus arrhythmia. *Biological Psychology*, *74*(2), 286–294. <https://doi.org/10.1016/j.biopsycho.2005.09.005>
- Duschek, S., Muckenthaler, M., Werner, N., & Reyes Del Paso, G. A. (2009). Relationships between features of autonomic cardiovascular control and cognitive performance. *Biological Psychology*, *81*(2), 110–117. <https://doi.org/10.1016/j.biopsycho.2009.03.003>
- Evans, A. M., Dillon, K. D., & Rand, D. G. (2015). Fast but not intuitive, slow but not reflective: Decision conflict drives reaction times in social dilemmas. *Journal of Experimental Psychology: General*, *144*(5), 951–966. <https://doi.org/10.1037/xge0000107>
- Evans, J. S. B. T., & Stanovich, K. E. (2013). Dual-process theories of higher cognition: Advancing the debate. *Perspectives on Psychological Science*, *8*(3), 223–241. <https://doi.org/10.1177/1745691612460685>
- Forste, G., Favieri, F., & Casagrande, M. (2019). Heart rate variability and cognitive function: A systematic review. *Frontiers in Neuroscience*, *13*, 1–11. <https://doi.org/10.3389/fnins.2019.00710>
- Friedman, B. H. (2007). An autonomic flexibility-neurovisceral integration model of anxiety and cardiac vagal tone. *Biological Psychology*, *74*(2), 185–199. <https://doi.org/10.1016/j.biopsycho.2005.08.009>
- Gratz, K., & Roemer, L. (2004). Multidimensional assessment of emotion regulation and dysregulation: Development, factor structure, and initial validation of the difficulties in emotion regulation scale. *Journal of Psychopathology and Behavioral Assessment*, *26*(1), 41–54. <https://doi.org/10.1023/B:JOBA.0000007455.08539.94>
- Greene, J. D., Morelli, S. A., Lowenberg, K., Nystrom, L. E., & Cohen, J. D. (2008). Cognitive load selectively interferes with utilitarian moral judgment. *Cognition*, *107*(3), 1144–1154. <https://doi.org/10.1016/j.cognition.2007.11.004>
- 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. <https://doi.org/10.1016/j.neuron.2004.09.027>
- 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. <https://doi.org/10.1126/science.1062872>
- Grossman, P., Karemaker, J., & Wieling, W. (1991). Prediction of tonic parasympathetic cardiac control using respiratory sinus arrhythmia: The need for respiratory control. *Psychophysiology*, *28*(2), 201–216. <https://doi.org/10.1111/j.1469-8986.1991.tb00412.x>
- Gürçay, B., & Baron, J. (2017). Challenges for the sequential two-system model of moral judgement. *Thinking & Reasoning*, *23*(1), 49–80. <https://doi.org/10.1080/13546783.2016.1216011>
- Hansen, A. L., Johnsen, B. H., & Thayer, J. F. (2003). Vagal influence on working memory and attention. *International Journal of Psychophysiology*, *48*(3), 263–274. [https://doi.org/10.1016/S0167-8760\(03\)00073-4](https://doi.org/10.1016/S0167-8760(03)00073-4)
- Holzman, J. B., & Bridgett, D. J. (2017). Heart rate variability indices as bio-markers of top-down self-regulatory mechanisms: A meta-analytic review. *Neuroscience and Biobehavioral Reviews*, *74*, 233–255. <https://doi.org/10.1016/j.neubiorev.2016.12.032>
- Huebner, B., Hauser, M. D., & Pettit, P. (2011). How the source, inevitability and means of bringing about harm interact in folk-moral judgments. *Mind & Language*, *26*(2), 210–233.
- Hutcherson, C. A., Montaser-Kouhsari, L., Woodward, J., & Rangel, A. (2015). Emotional and utilitarian appraisals of moral dilemmas are encoded in separate areas and integrated in ventromedial prefrontal cortex. *The Journal of Neuroscience*, *35*(36), 12593–12605. <https://doi.org/10.1523/JNEUROSCI.3402-14.2015>
- Ingjaldsson, J. T., Laberg, J. C., & Thayer, J. F. (2003). Reduced heart rate variability in chronic alcohol abuse: Relationship

- with negative mood, chronic thought suppression, and compulsive drinking. *Biological Psychiatry*, 54(12), 1427–1436. [https://doi.org/10.1016/S0006-3223\(02\)01926-1](https://doi.org/10.1016/S0006-3223(02)01926-1)
- Jennings, J., Allen, B., Gianaros, P., Thayer, J., & Manuck, S. (2015). Focusing neurovisceral integration: Cognition, heart rate variability, and cerebral blood flow. *Psychophysiology*, 52(2), 214–224. <https://doi.org/10.1111/psyp.12319>
- Kok, B. E., & Fredrickson, B. L. (2010). Upward spirals of the heart: Autonomic flexibility, as indexed by vagal tone, reciprocally and prospectively predicts positive emotions and social connectedness. *Biological Psychology*, 85(3), 432–436. <https://doi.org/10.1016/j.biopsycho.2010.09.005>
- Krajbich, I., Bartling, B., Hare, T., & Fehr, E. (2015). Rethinking fast and slow based on a critique of reaction time reverse inference. *Nature Communications*, 6(1), 7455. <https://doi.org/10.1038/ncomms8455>
- Luft, C. D. B., Takase, E., & Darby, D. (2009). Heart rate variability and cognitive function: Effects of physical effort. *Biological Psychology*, 82(2), 186–191. <https://doi.org/10.1016/j.biopsycho.2009.07.007>
- McIntosh, R. C., Hoshi, R., Nomi, J. S., Di Bello, M., Goodman, Z. T., Kornfeld, S., Uddin, L. Q., & Ottaviani, C. (2020). Neurovisceral integration in the executive control network: A resting state analysis. *Biological Psychology*, 157, 107986. <https://doi.org/10.1016/j.biopsycho.2020.107986>
- Middleton, H. C., Sharma, A., Agouzoul, D., Sahakian, B. J., & Robbins, T. W. (1999). Contrasts between the cardiovascular concomitants of tests of planning and attention. *Psychophysiology*, 36(5), 610–618. <https://doi.org/10.1111/1469-8986.3650610>
- Miller, E. K., & Cohen, J. D. (2001). An integrative theory of prefrontal cortex function. *Annual Review of Neuroscience*, 24, 167–202.
- 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. <https://doi.org/10.1111/j.1467-9280.2008.02122.x>
- Mulcahy, J. S., Larsson, D. E., Garfinkel, S. N., & Critchley, H. D. (2019). Heart rate variability as a biomarker in health and affective disorders: A perspective on neuroimaging studies. *Neuroimage*, 202, 116072. <https://doi.org/10.1016/j.neuroimage.2019.116072>
- Muthukrishnan, S. P., Gurja, J. P., & Sharma, R. (2017). Does heart rate variability predict human cognitive performance at higher memory loads? *Indian Journal of Physiology and Pharmacology*, 61(1), 14–22. https://ijpp.com/IJPP%20archives/2017_61_1/14-22.pdf
- Park, G., Kappes, A., Rho, Y., & Van Bavel, J. J. (2016). At the heart of morality lies neuro-visceral integration: Lower cardiac vagal tone predicts utilitarian moral judgment. *Social Cognitive and Affective Neuroscience*, 11(10), 1588–1596. <https://doi.org/10.1093/scan/nsw077>
- Park, G., Vasey, M. W., Van Bavel, J. J., & Thayer, J. F. (2013). Cardiac vagal tone is correlated with selective attention to neutral distractors under load. *Psychophysiology*, 50(4), 398–406. <https://doi.org/10.1111/psyp.12029>
- Park, G., Vasey, M. W., Van Bavel, J. J., & Thayer, J. F. (2014). When tonic cardiac vagal tone predicts changes in phasic vagal tone: The role of fear and perceptual load. *Psychophysiology*, 51(5), 419–426. <https://doi.org/10.1111/psyp.12186>
- Porges, S. W. (1995). Orienting in a defensive world: Mammalian modifications of our evolutionary heritage. A polyvagal theory. *Psychophysiology*, 32(4), 301–318. <https://doi.org/10.1111/j.1469-8986.1995.tb01213.x>
- Rosas, A., & Aguilar-Pardo, D. (2020). Extreme time-pressure reveals utilitarian intuitions in sacrificial dilemmas. *Thinking & Reasoning*, 26(4), 534–551. <https://doi.org/10.1080/13546783.2019.1679665>
- Rosas, A., Bermúdez, J. P., & Aguilar-Pardo, D. (2019). Decision conflict drives reaction times and Utilitarian responses in sacrificial dilemmas. *Judgment and Decision Making*, 14(5), 555–564. <http://journal.sjdm.org/19/190615/jdm190615.pdf>
- Rosas, A., & Koenigs, M. (2014). Beyond “utilitarianism”: Maximizing the clinical impact of moral judgment research. *Social Neuroscience*, 9(6), 661–667. <https://doi.org/10.1080/17470919.2014.937506>
- Segerstrom, S. C., & Nes, L. S. (2007). Heart rate variability reflects self-regulatory strength, effort, and fatigue. *Psychological Science*, 18(3), 275–281. <https://doi.org/10.1111/j.1467-9280.2007.01888.x>
- Shenhav, A., & Greene, J. D. (2014). Integrative moral judgment: Dissociating the roles of the amygdala and ventromedial prefrontal cortex. *Journal of Neuroscience*, 34(13), 4741–4749. <https://doi.org/10.1523/JNEUROSCI.3390-13.2014>
- Smith, R., Thayer, J. F., Khalsa, S. S., & Lane, R. D. (2017). The hierarchical basis of neurovisceral integration. *Neuroscience & Biobehavioral Reviews*, 75, 274–296. <https://doi.org/10.1016/j.neubiorev.2017.02.003>
- Smith, T. W., Cribbet, M. R., Nealey-Moore, J. B., Uchino, B. N., Williams, P. G., MacKenzie, J., & Thayer, J. F. (2011). Matters of the variable heart: Respiratory sinus arrhythmia response to marital interaction and associations with marital quality. *Journal of Personality and Social Psychology*, 100(1), 103–119. <https://doi.org/10.1037/a0021136>
- Sosnowski, M. (2010). Heart rate variability. In P. W. Macfarlane, A. van Oosterom, O. Pahlm, P. Kligfield, M. Janse, J. Camm (Eds), *Comprehensive electrocardiology* (pp. 1513–1675). Springer.
- Tangney, J. P., Baumeister, R. F., & Boone, A. L. (2004). High self-control predicts good adjustment, less pathology, better grades, and interpersonal success. *Journal of Personality*, 72(2), 271–324. <https://doi.org/10.1111/j.0022-3506.2004.00263.x>
- Task Force of The European Society of Cardiology and The North American Society of Pacing and Electrophysiology. (1996). Heart rate variability: Standards of measurement, physiological interpretation and clinical use. Task Force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology. *Circulation*, 93(5), 1043–1065. <https://doi.org/10.1161/01.CIR.93.5.1043>
- Thayer, J. F., Hansen, A. L., & Johnsen, B. H. (2010). The non-invasive assessment of autonomic influences on the heart using impedance cardiography and heart rate variability. In A. Steptoe (Ed.), *Handbook of Behavioral Medicine* (pp. 723–740). Springer.
- Thayer, J. F., Hansen, A. L., Saus-Rose, E., & Johnsen, B. H. (2009). Heart rate variability, prefrontal neural function, and cognitive performance: The neurovisceral integration perspective on self-regulation, adaptation, and health. *Annals of Behavioral Medicine*, 37(2), 141–153. <https://doi.org/10.1007/s12160-009-9101-z>

- Thayer, J. F., & Lane, R. D. (2000). A model of neurovisceral integration in emotion regulation and dysregulation. *Journal of Affective Disorders*, 61(3), 201–216. [https://doi.org/10.1016/S0165-0327\(00\)00338-4](https://doi.org/10.1016/S0165-0327(00)00338-4)
- Trémolière, B., & Bonnefon, J.-F. (2014). Efficient kill–save ratios ease up the cognitive demands on counterintuitive moral utilitarianism. *Personality and Social Psychology Bulletin*, 40(7), 923–930. <https://doi.org/10.1177/0146167214530436>
- Weber, C. S., Thayer, J. F., Rudat, M., Wirtz, P. H., Zimmermann-Viehoff, F., Thomas, A., ... Deter, H. C. (2010). Low vagal tone is associated with impaired post stress recovery of cardiovascular, endocrine, and immune markers. *European Journal of Applied Physiology*, 109(2), 201–211. <https://doi.org/10.1007/s00421-009-1341-x>
- Zahn, D., Adams, J., Krohn, J., Wenzel, M., Mann, C. G., Gomille, L. K., Jacobi-Scherbening, V., & Kubiak, T. (2016). Heart rate variability and self-control—A meta-analysis. *Biological Psychology*, 115, 9–26. <https://doi.org/10.1016/j.biopsycho.2015.12.007>
- Bago, B., & De Neys, W. (2019). The intuitive greater good: Testing the corrective dual process model of moral cognition. *Journal of Experimental Psychology: General*, 148(10), 1782–1801. Advance online publication. <https://doi.org/10.1037/xge0000533>

Appendix

English translations and original scenarios (Spanish).

The utilitarian incentives built into each scenario are indicated in parentheses in the English version.

Shark (baseline incentive: save five)

You are watching an exhibition of sharks being fed in an aquarium pool. A metal fence suddenly collapses, and a group of people fall into the pool. Their frantic movement in the water attracts the hungry sharks. A person next to you has a harmless episode of nose-bleeding. If you push this person into the water, their blood will draw the sharks away from the other five people. The bleeding person will die, but the five people will swim to safety.

How right or wrong is it to cause the death of the person in order to save five people?

Boat (baseline and doomed victim)

You escape on a lifeboat from a ship on fire. The lifeboat slowly sinks due to the weight of 2 passengers too many. You are a good swimmer, so you jump into the turbulent waters; but now with one passenger too many the boat continues to slowly sink. None of the remaining passengers wear life jackets, and none of them can swim.

Nobody is willing to get off and the boat will soon sink. If you pull the person closest to you into the water and keep clear of his attempts to drag you down with him, this person will drown, but you will save five people in the boat.

How right or wrong is it to cause the death of the person in order to save five others?

Dam (baseline and save100K)

You have a job in the dam management of a large river. Failure in the materials has caused a leak in the dam. The leak has opened a hole and the dam will soon burst apart. A huge volume of water will flood a nearby town. There is no time to warn the inhabitants.

On the dam's edge a worker looks concerned into the dam. If you push this person into the water, his body will stopple the hole, giving time to take the necessary actions. This person will drown, but you will save 100.000 people living in the town.

How right or wrong is it to cause the death of the person in order to save 100.000 people?

Flames (baseline and selfish incentive)

You and other five people are trapped in a burning building. There is only one emergency exit, but it is blocked by burning debris. A sixth person tries to crawl through a hole at the base of that exit. You and the other five lack the time to do the same thing.

If you push that person through the debris to unlock the exit, that person will die, but you and the other five people will exit to safety.

How right or wrong is it to cause the death of the person in order to save yourself and four others?

Grenade (baseline and guilty victim)

You stand on a rooftop and near you a man is about to throw a grenade onto a group of five people below. The group is unaware of the threat and there is no way to warn them. The five people will surely die in the explosion.

If you push the man with the grenade with a quick movement, he will fall and fail to activate the grenade. The fall will kill him, but the five people below will be saved.

How right or wrong is it to cause the death of the person in order to save five others?

Original scenarios (Spanish)

Tiburón

Visitas un acuario que exhibe tiburones mientras son alimentados. Muy cerca de ti, la reja donde se agolpan los visitantes cede y 5 personas caen al agua. Su caída atrae a los tiburones hambrientos.

Junto a ti, otro visitante sangra excesivamente por la nariz. Si lo empujas al agua, la sangre atraerá a los tiburones y los alejará de las 5 personas que están a punto de ser devoradas. El hombre que sangra morirá, pero las otras cinco personas se salvarán.

Barco

Viajas en un barco que se incendia y te pones a salvo en un bote. El bote aguanta a cinco personas, pero lleva siete y comienza a hundirse por sobrepeso. Tú nadas muy bien y saltas al agua, pero el bote con seis se sigue hundiendo. Ninguno de los seis lleva chaleco salvavidas y ninguno sabe nadar. Todos van a morir si nadie salta al agua; pero nadie lo hace.

Si desde el agua tú jalas al hombre que está más cerca y lo sacas del bote sabiendo que se ahogará por las condiciones difíciles del mar, el bote dejará de hundirse y los otros cinco se salvarán.

Represa

Trabajas en la represa de un río muy caudaloso. Por fallas en el material se produjo un escape en el dique. La concentración de la presión en ese punto romperá el dique. El río caerá con ferocidad sobre una ciudad aledaña de cien mil habitantes. Ya no hay tiempo de avisar.

En el borde de la represa hay un trabajador. Si lo empujas al agua, la corriente lo arrastrará y tamará el hueco con su cuerpo dando tiempo a tomar medidas permanentes. El trabajador se ahogará, pero se salvarán cien mil vidas.

Llamas

Tú y cuatro personas más están atrapadas en un edificio en llamas. Hay una única salida de emergencia, pero está bloqueada por escombros ardientes. Un hombre intenta arrastrarse por un hueco en la base de dicha salida, pero las otras cuatro personas y tú no tienen tiempo de hacer lo mismo.

Si empujas al hombre que está intentando salir contra los escombros ardientes para desbloquear el paso, él morirá, pero tú y las otras cuatro personas se salvarán.

Granada

Estás en una azotea y te das cuenta de que cerca de ti hay un hombre que amenaza con arrojar una granada sobre un grupo de 5 personas reunidas en un parque aledaño. El grupo desconoce la amenaza y no hay forma de avisarles. Las 5 personas morirán con seguridad en la explosión.

Si empujas al hombre con un movimiento rápido, caerá desde la azotea sin tener tiempo de activar la granada. Esto lo matará, pero así se salvarán las 5 personas inocentes.