Moral judgment mechanisms reflected in brainwaves: An ERP study

J. Meng[1], X. Xu[2], Y. Zhong[2], S. Chandler[1], M. F. Curado[1], C. Kuntz[1], O. R. Shaw[1], D. Zhang[1], E. Kaan[1]

[1]University of Florida, [2]Nanjing Normal University

Moral judgment has been viewed as consisting of two underlying mechanisms: affective and normative (Nichols, 2002). While the affective process is often associated with an initial emotional reaction to the moral stimuli, the normative mechanism reflects a slower evaluative process. Previous ERP studies on moral judgment have identified components that might reflect these two processes. Early ERP components such as P200 and N2, typically ^{obs2erved} in the frontal regions, appear to correspond to initial affective reactions (Lu et al., 2019). Late positive components observed in the parietal regions (LPP: Late Posterior Positivity; LPC: Late Positivity Component), seem to correspond to the later evaluative moral decisions (Kunkel et al., 2018). However, some have questioned whether moral violations are processed differently from semantic violations. The ERP components observed in moral violation conditions could be an effect of lower plausibility (Leuthold et al., 2015). So far, few ERP studies have directly compared moral and semantic violations. The current study investigates English speakers' processing of moral violations contained in sentences, and how this differs from the processing of semantic violations.

Experimental stimuli were in English and consisted of three conditions: moral violation, semantic violation, and a neutral condition (see Table 1). Items were normed for plausibility (neutral: M=4.70, moral: M=2.73, semantic: M=1.63) and moral acceptability (neutral: M =3.72, moral: M =1.62, semantic M = 2.81). Each trial started with a context sentence, different for each condition, presented in its entirely. This was followed by a target sentence containing words that were neutral or induced a moral or semantic violation given the context. The target sentence was the same for each condition and was presented word by word. 37 native English-speaking young adults (17 f, 10m, 1 non-binary) silently read 24 items per condition while their EEG was recorded. Comprehension questions randomly appeared after 53.3% trials.

Mean amplitude was obtained from three time windows at the critical word region: early potentials from 150ms to 250ms (frontocentral electrodes), N400 from 300ms to 500ms (central electrodes), and late potentials from 500ms to 800ms (central-parietal electrodes). Linear mixed effects models were used to analyze the mean amplitudes, using condition (Helmert-coded) as fixed effects, and a by-subject random intercept. In the 500-800ms window, the two violation conditions showed a significantly larger LPC amplitude than the neutral condition [b=.74, 95%CI [.08,1.40], SE=.33; T=2.24, $p < .05$]. The moral condition had a more positive amplitude than the semantic condition [b=1.12, 95%CI[.36,1.88], SE=.38; T=2.93, $p < .01$]. The latter effect already started in the 300-500 ms N400 window [b=1.05, 95%CI [.30,1.79], SE=.38; T=2.79, $p < .01$] (see Figure 1). The LPC had a more posterior distribution for the moral than the semantic violation (see Figure 2). No significant effects were found for the early potentials. The LPC effects might reflect a later evaluative process after the participants received information that was not congruent with their socio-normative knowledge. The earlier onset, larger amplitude, and more posterior distribution of this effect for moral violations suggest that moral information is processed differently from semantic information. We plan on conducting further analysis will be done to see how individuals' moral attitudes and ratings are associated with their ERPs to moral violations.

Table 1. Experimental Conditions.

Conditions	Context	Target sentence
Moral	As Rose was driving to work, some school children started crossing the road in front of her.	So| she| **sped up**| the car| immediately.
Semantic	Rose was preparing to go to work today but discovered that her car had no gas.	So| she| **sped up**| the car| immediately.
Neutral	As Rose was driving to work, she realized she was going to be late.	So| she| **sped up**| the car| immediately.

Note. Critical word region is in bold.

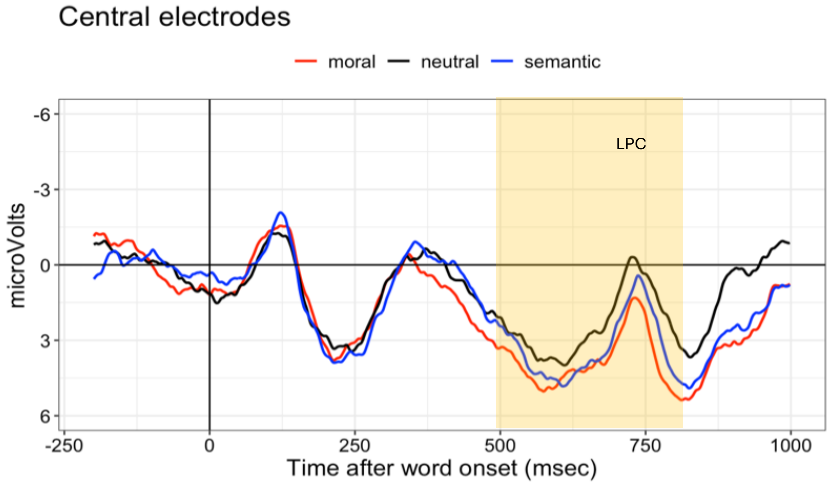

Figure 1. ERPs at critical word at the Cz electrode. Black: neutral condition; red: moral violation; blue: semantic violation. LPC time window (500-800 ms) is in yellow.

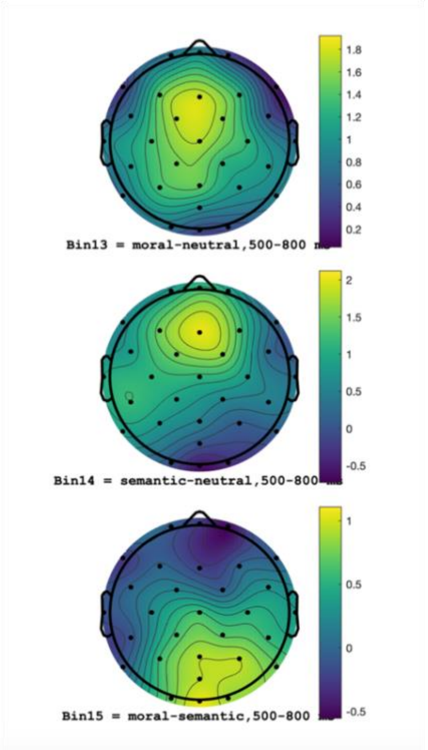

Figure 2. Topographic plots of difference waves (500-800 ms). Top: Moral minus neutral; Middle: Semantic minus neutral; Bottom: Moral minus semantic. Yellow indicates a more positive difference. The LPC had a more posterior distribution for the moral than the semantic violation conditions ERPs.

References: Kunkel, A., Filik, R., Mackenzie, I. G., & Leuthold, H. (2018). Task-dependent evaluative processing of moral and emotional content during comprehension: An ERP study. *Cognitive, Affective, & Behavioral Neuroscience*, *18*(2), 389–409. Leuthold, H., Kunkel, A., Mackenzie, I. G., & Filik, R. (2015). Online processing of moral transgressions: ERP evidence for spontaneous evaluation. *Social Cognitive and Affective Neuroscience*, *10*(8), 1021–1029. Lu, J., Peng, X., Liao, C., & Cui, F. (2019). The stereotype of professional roles influences neural responses to moral transgressions: ERP evidence. *Biological Psychology*, *145*, 55–61. Nichols, S. (2002). Norms with feeling: Towards a psychological account of moral judgment. *Cognition*, *84*(2), 221–236.