Supplementary Material

Examining the Effect of Time-on-task on the ERN

We ran parallel analyses examining change in ERN amplitude as a function of time-on-task, rather than ordered errors as reported in the main manuscript. As in the main analyses, subjects and electrodes were included as crossed random factors. The intercept and slopes of race and object were allowed to vary by subject; the intercept was also allowed to vary by electrode. The dependent variable was mean ERN amplitude quantified following each error committed by each subject (i.e., trial-level data). The original trial number for each error was maintained so that time-on-task was maintained between errors (i.e., an error committed on trial 10 was not equivalent to an error committed on trial 50, even if they were both the first error committed by a participant). Race of the prime (Black, White), the type of object (gun, tool), and error number were included as level-1 predictors of ERN amplitude. Fixed effects of race and object are interpreted at the first trial.

This model estimated a significant effect of race, $b = .56$, $p < .001$, object, $b = -.66$, $p < .001$, and the race x object interaction, $b = .34$, $p < .001$. These effects were all in the same direction as those found when using ordered errors rather than time-on-task as the continuous predictor. Additionally, the effect of trial was significant, $b = .07$, $p < .001$, such that ERN amplitude became more positive (smaller) as the task progressed. This main effect was qualified by a trial x object interaction, $b = .03$, $p = .004$, such that the positive association between trial and ERN amplitude was stronger for errors following tools than errors following guns. Lastly, the three way trial x race x object interaction was significant, $b = -.02$, $p = .010$. Examination of the simple slopes revealed a positive slope for ERNs following all types of errors except those

categorizing guns following Black faces. Although this three-way interaction was only marginally significant when error number was included in the model, the pattern is the same.

Subject and Electrode ICCs

As can be seen in Table 1, the ICC associated with subject is much larger when using mean amplitudes calculated from averaged waveforms as the DV (i.e., the signal averaging approach) than when using mean amplitudes calculated from each trial separately as the DV (i.e., the trial level approach). As indicated by the variances, the difference in ICCs is due to a much larger residual variance in the trial level data compared to the averaged waveforms, rather than any differences in subject or electrode variance, which are similar across the two approaches. This larger residual variance contributes to a larger denominator in the calculation for ICC and results in a smaller ICC for both subject and electrode.

The much smaller residual variance in the signal averaging approach is due to the observations included in the model, as the DVs are essentially sample means (mean amplitudes quantified from waveforms created by averaging across a number of trials). The more trials included in the average, the more noise is eliminated from the sample mean, and the less residual variance remains in the DV. In contrast, the noise in the raw trial level data is included in the DV in the trial level approach, resulting in a much larger residual variance. This pattern parallels the central limit theorem's prediction that sample means will have less variability than raw data, and that the larger the number of samples included in each sample mean, the less variability in sample means.

Given that much of the within-subject variance has been removed in the signal averaging approach, it is perhaps unsurprising that subject ICCs in the signal averaging approach are quite large (.59 in Study 1 and .48 in Study 2), indicating that the majority or close to the majority of

variance in (averaged) signal comes from between-subject differences. This parallels previous work reporting that ERPs elicited by faces are highly reproducible within subjects, but differ reliably between subjects (ICCs ranged from .3 to .9 across time points; Gaspar, Rousselet, & Pernet, 2011). Although Gaspar et al. (2011) did not directly test the sources of individual differences in ERPs, they speculate that differences can arise from both trivial (head size, skull thickness, etc.) and non-trial factors (brain anatomy, variation in information processing)

Compared to the between-subjects variability, there was much less between-electrode variability. Low variability may arise from pre-selection of electrodes at which the component of interest is most apparent. In other words, if all the recorded electrodes were included in the analysis, the electrode-ICC would likely be much higher due to variation in the extent to which the component is evident at each electrode.

Supplementary Tables

Table S1

Random Effects of Multilevel Model Testing the Effect of Race and Trial on the Mean Amplitude of the N170.

Groups		Variance	Std. Dev	Corr.
Subject	Intercept	4.56*	2.14	
	Race	0.35*	0.60	-0.19
Electrode	Intercept	2.39*	1.55	
Residual		100.06	10.00	

Note. Models estimating the effect of race at the beginning and end of the experiment have the same random effects. Variances of random intercepts and slopes were tested using likelihood ratio tests. Asterisks indicate a significant chi square value.

Table S2

Random Effects of Multilevel Model Testing the Effect of Race, Object and Trial on the Mean Amplitude of the ERN.

Groups		Variance	Std. Dev	Corr.		
Subject	Intercept	8.24*	2.87			
	Race	0.91*	0.96	-0.10		
	Object	1.58*	1.26	-0.02	-0.01	
	Race x Object	0.78*	0.88	0.00	-0.20	-0.06
Electrode	Intercept	2.02*	1.42			
Residual		57.73	7.60			

Note. Corresponds to model presented in Table 2. Variances of random intercepts and slopes were tested using likelihood ratio tests. Asterisks indicate a significant chi square value.