*Results*

Data file opened 1/13/20 2:53pm EST

First 750 analyzed 1/13/20 3:25pm EST

Second 750 analyzed 1/13/20 3:30pm EST

Full Sample Analyzed 1/13/20 3:34pm EST

The Luth survey company recruited participants to take part in the study. We set a target sample size of 1,500 participants, to be collected in two waves of 750. Luth collects data from participants in 20-minute sessions and considers a complete survey to be one in which the participant finishes the entire 20-minute session. Because each session contains a number of studies, there is inevitably dropout between the first and last study in the session; this means that although ~1500 participants completed the total session, 2,388 participants consented to complete the target study (“Redemption), completed the study, and did not use a mobile device to complete the study (participants were required to complete the device on a desktop computer). Of these participants, 2,117 successfully answered our pre-registered attention check item. Similarly, although we refer to participants being split into two waves of 750, more than 750 participants completed the target study in each wave. The waves of the study are identified by when data collection began (9/9/19 for wave 1 and 9/13/19 at 4:45pm EST for wave 2.)

Participants read that they were providing their impressions of recent criminal offenders. Each participant provided their impressions of 4 offenders. The evaluated offenders were counterbalanced by gender and their offence. The primary dependent variable was the participants’ averaged responses across the four offenders to the question “How remorseful does \_\_\_\_ seem?”. We created a dummy variable for condition, which coded such that 1 = participants saw an appearance change and 0 = if the participants did not see an appearance change.

**Wave 1 Results**

1,177 Wave 1 participants completed the target study and correctly responded to the pre-registered attention check items.

We conducted an independent sample *t­­*-test to compare participants averaged responses to the “remorse” item between conditions. We found that participants evaluated offenders who had undergone an appearance change to be more remorseful than those who did not, *M*change = 3.98, *SD* = 0.93, *N* = 593; *M*control = 3.86, *SD* = 0.97, *N* = 584; *t*(1175) = 2.27, *p* = .023, Cohen’s *d* = .133.

**Wave 2 Results**

940 Wave 2 participants completed the target study and correctly responded to the pre-registered attention check items.

We conducted an independent sample *t­­*-test to compare participants averaged responses to the “remorse” item between conditions. We found that participants evaluated offenders who had undergone an appearance change to be more remorseful than those who did not, *M*change = 4.00, *SD* = 0.89, *N* = 472; *M*control = 3.87, *SD* = 0.98, *N* = 468; *t*(938) = 2.16, *p* = .031, Cohen’s *d* = .141.

**Total Sample Results**

2,117 participants completed the target study and correctly responded to the pre-registered attention check items.

We conducted an independent sample *t­­*-test to compare participants averaged responses to the “remorse” item between conditions. We found that participants evaluated offenders who had undergone an appearance change to be more remorseful than those who did not, *M*change = 3.99, *SD* = 0.91, *N* = 1,065; *M*control = 3.86, *SD* = 0.97, *N* = 1,052; *t*(2115) = 3.13, *p* = .002, Cohen’s *d* = .136.