**Data Analysis Plan**

**Data Preparation**

**Self-reported ratings**. A mean self-reported rating score will be calculated for Chris by averaging responses from the three Likert rating scales. Positive values indicate positive evaluations of Chris, negative values indicate negative evaluations, and neutral evaluations indicate neutral or ambivalent responses to Chris.

**IAT**. The D2 algorithm will be used to create pIAT scores. Scores will be calculated so that positive values reflected a relative response bias for Chris over Bob whereas negative values indicated the reverse pattern of responding (a relative response bias favoring Bob over Chris).

**Behavioral intentions.**

**Data Exclusions**

1. The data of participants who do not fully complete all questions and tasks will be excluded from analyses during the first round of analyses.

2. The data will be excluded of those participants who had pIAT error rates for any of the pIATs above 30% across the entire task, or above 40% for any one of the four critical blocks or for participants who complete more than 10% of pIAT trials faster than 400 ms.

**Hypothesis Testing**

**Confirmatory Analysis 1a: Presence of evaluations and intentions following genuine videos.**

This mean score will be submitted to single sample t-tests to examine if self-reported and pIAT scores differ from zero, one for those in the positive content condition and another for those in the negative content condition.

**Confirmatory Analysis 1b: Moderation of evaluations and intentions by video content**. An average self-reported rating score for Chris will be calculated by averaging responses from the three Likert rating scales. This mean score will be submitted to an independent samples t-test with *video content* (Positive vs. Negative) as a between subjects factor. pIAT scores will be submitted to a similar set of analyses.

In all cases, effect sizes (Cohen’s d) will be reported. We will also compute Bayesian factors in accordance with procedures outlined by Rouder, Speckman, Sun, Morey, and Iverson (2009) to estimate the amount of evidence for the hypothesis that stimulus evaluations differ as a function of video content (alternative hypothesis) or that there is no difference (null hypothesis).

**Moderation of evaluations by video type**. An independent samples t-test will be carried out in order to examine if the genuine and Deepfaked videos differ in the evaluations that they produce. Data will first be recoded so that the valence of the video content is controlled for (i.e., scores from those in the negative content groups will be re-coded by multiplying their values by -1). Effect sizes (Cohen’s d) will be reported. We will also compute Bayesian factors in accordance with procedures outlined by Rouder, Speckman, Sun, Morey, and Iverson (2009) to estimate the amount of evidence that stimulus evaluations differ as a function of video type (alternative hypothesis) or that there is no difference (null hypothesis).

**Moderation of evaluations by Deepfake detection**. An independent samples t-test will be carried out in order to examine if Deepfake detectors (i.e., those that indicate “Yes” on the deepfake awareness question) and Deepfake non-detectors (i.e., those that indicate “No” on the same question) differ in the evaluations that they produce. Data will first be recoded so that the valence of the video content is controlled for (i.e., scores from those in the negative content groups will be re-coded by multiplying their values by -1). Effect sizes (Cohen’s d) will be reported. We will also compute Bayesian factors in accordance with procedures outlined by Rouder, Speckman, Sun, Morey, and Iverson (2009) to estimate the amount of evidence that stimulus evaluations differ as a function of deepfake detection (alternative hypothesis) or that there is no difference (null hypothesis).