**Supplemental Materials**

**An Illusion of Unfairness About Random Outcomes**

Contents

[Table S1: List of All Studies and Additional Independent Variables 4](#_Toc187657383)

[Preregistrations 8](#_Toc187657384)

[Outcome Variable Names Across the 11 Studies 9](#_Toc187657385)

[Data Inclusion 10](#_Toc187657386)

[Power Analyses 10](#_Toc187657387)

[*t*-test Results for the Illusion of Unfairness Across 9 Studies 16](#_Toc187657388)

[Table S2. Summary of *t*-test results for the basic illusion of unfairness effect (9 Studies). 17](#_Toc187657389)

[Table S3. Summary of *t*-test results for the downstream effects due to the illusion of unfairness effect (9 studies). 19](#_Toc187657390)

[Meta-Analysis of the Illusion of Unfairness Results: Extended Methods and Results 21](#_Toc187657391)

[Table S4a. Meta-Analytic Effects of the Illusion of Unfairness Across 9 “Losing” Studies 22](#_Toc187657392)

[Table S4b. Meta-Analytic Effects of the Illusion of Unfairness Across 3 ‘Winning’ Studies 23](#_Toc187657393)

[Study 1 25](#_Toc187657394)

[Table S5. Study 1 Descriptives and T-Tests 29](#_Toc187657395)

[29](#_Toc187657396)

[Table S6. Experimenter Flip vs P-Flipped vs Other-P Flipped Contrasts 32](#_Toc187657397)

[Study 2 33](#_Toc187657398)

[Table S7. *t*-test results Study 2, Lose Condition 35](#_Toc187657399)

[Table S8. *t*-test results Study 2, Win Condition 37](#_Toc187657400)

[Study 3 38](#_Toc187657401)

[Table S9. Descriptive and *t*-test Results of Study 3 41](#_Toc187657402)

[Study 4 46](#_Toc187657403)

[Table S10. Descriptives and *t*-test results Study 4 (winning outcome) 46](#_Toc187657404)

[Studies 5 and 6 49](#_Toc187657405)

[Table S11. Descriptive and t-test Results of Study 5 54](#_Toc187657406)

[Table S12. Descriptives and *t*-test Results of Study 6 56](#_Toc187657407)

[Study 7 58](#_Toc187657408)

[Table S13: Descriptives and t-test results Study 7 60](#_Toc187657409)

[Study 8 64](#_Toc187657410)

[Table S14: Study 8: Effects of Who Flipped and Consenting to How it Would Be Determined Who Flipped 67](#_Toc187657411)

[Table S15: T-Test Results of Study 8 70](#_Toc187657412)

[Study 9 72](#_Toc187657413)

[Table S16. Descriptives and *t*-test Results at Time 2 for Study 9 75](#_Toc187657414)

[Study 10 77](#_Toc187657415)

[Table S17. Descriptives and *t*-test Results Study 10 79](#_Toc187657416)

[Table S18. Results Comparing Explanation Condition with No Explanation Condition Using only Other-P Flipped Condition 82](#_Toc187657417)

[Table S19. Results Comparing Explanation Condition with No Explanation Condition using Only Participant Flip Condition 83](#_Toc187657418)

[Study 11 84](#_Toc187657419)

[Study 12 87](#_Toc187657420)

[Table S20. Descriptives and *t*-test results Study 12 (Lose Condition) 89](#_Toc187657421)

[Table S21. Descriptives and *t*-test results Study 12 (Win Condition) 91](#_Toc187657422)

[Moral Luck 93](#_Toc187657423)

[Correlation Matrices between Outcome Variables in All Studies 94](#_Toc187657424)

# Table S1: List of All Studies and Additional Independent Variables

|  |  |  |  |
| --- | --- | --- | --- |
| **Study** | **Setting** | **Additional Independent Variables (other than the main variable of who called heads/tails and flipped the coin)** | **Results of Additional Condition(s)** |
| 1 | Lab | Included a third condition in which the Experimenter Flipped the Coin | In the experimenter flipped condition, participants’ responses were generally in between the responses of those in the P-Flipped and Other-P Flipped condition; see Table S6. |
| 2 | Lab | Two participants competed, with one losing the coin flip and the other winning the flip. | Reported in Paper |
| 3 | Field | Two participants competed, with one losing the coin flip (reported in the meta-analysis) and the other winning the flip | Reported in Paper |
| 4 | Field | Participants competed against a confederate and always won the coin flip | Reported in Paper |
| 5 | mTurk | Participants learned that their opponent was either White or Black, to see if the illusion of unfairness is increased when one’s opponent is of a different race. | There were very few significant main effects of race of the other participant or Race x Who Flipped interactions, no more than would be expected by chance. We thus collapsed across this variable. |
| 6 | mTurk | Participants learned that their opponent’s first name was Greg or Jamal (stereotypically White or Black names, respectively), to see if the illusion of unfairness is increased when one’s opponent is of a different race. | There were very few significant main effects of race of the other participant or Race x Who Flipped interactions, no more than would be expected by chance. We thus collapsed across this variable. |
| 7 | mTurk | We included two additional conditions that explored the effects of different ways of assigning participants to the negative outcome, other than a coin flip. In one condition, participants thought that the outcome was determined randomly in a way neither participant could control (who had more letters in their mTurk ID). In another, participants thought that the outcome was determined in a non-random manner, namely which participant responded most quickly to a prompt. Because neither condition involved perceived control over a negative outcome, they are not relevant to the current concerns. |  |
| 8 | mTurk | This study employed a 2 (Who Flipped: Participant or their Opponent) x Consent (no consent, consent) design. All participants were told that who got to call heads/tails and flip the coin was determined by the number of letters in their mTurk ID. In the consent condition, participants were asked to fill out a consent form acknowledging how random and arbitrary the coin flip assignment was. Participants in the no consent condition did not complete this form. | We replicated the illusion of unfairness on most measures. With one exception, asking participants to consent to how it was determined who would flip the coin did not significantly moderate this effect (See Table S13). The exception was on ratings of fairness. When no consent was given (as in previous studies), we replicated the illusion of unfairness: Participants thought it was less fair when the other participant got to flip the coin. When consent was given, who flipped did not influence fairness ratings. Why did participants in the consent condition moderate their responses on the fairness question? One possibility is experimental demand. Participants in the consent condition may have begrudgingly admitted that the procedure was fair, after agreeing with such statements as, “I understand it doesn't really matter who gets to call Heads/Tails and flip the coin, because the coin flip is random.” Nonetheless they still felt less responsible for the outcome when the other person flipped the coin, and were less pleased with the outcome (see Table S13). |
| 9 | mTurk | Participants completed the dependent measures twice, once before learning the outcome of the coin toss and again after learning the outcome. | Reported in Paper |
| 10 | mTurk | This study employed a 2 (Who Flipped: P-Flipped, Other-P Flipped) x 2 (Explanation: No Explanation, Random Explanation). The no explanation condition was the same as previous studies in which participants were given no rationale for who got to call heads/tails and flip the coin. In the explanation condition participants were told that who flipped would be determined by which participant had more letters in their mTurk ID, which is a random string of letters and numbers | Reported in Paper |
| 11 | mTurk | This is the one study in which we did not manipulate who flipped the coin. All participants learned that the other participant would do the flip. As described in the main paper, we manipulated how people were asked to respond to the dependent measures: with their “immediate gut reaction,” with their “logical, reasoned response,” or with no instructions. | Reported in Paper |
| 12 | Field | Participants competed in a coin flip against the experimenter to determine whether they would be compensated with a prize for filling out a survey. That is, instead of competing against another participant who was equally deserving of the reward, participants competed against the experimenter. If participants lost the flip they thus ended up helping the experimenter (by filling out a survey) for no compensation. If they won the flip they received appropriate compensation for their participation. Winning or losing was thus psychologically different than all the other studies, in which participants competed against another person of “equal standing.” (At the end of the study, after being debriefed, all participants received the prize.) | Reported in Supplements |

# Preregistrations

Study 1: no pre-registration

Study 2: <https://aspredicted.org/C6J_P1D>

Study 3: <https://aspredicted.org/5LS_42G>

Study 4: <https://aspredicted.org/SLD_XK4>

Study 5: <https://aspredicted.org/LSX_QW7>

Study 6: <https://aspredicted.org/46L_DHD>

Study 7: <https://aspredicted.org/LH6_PCV>

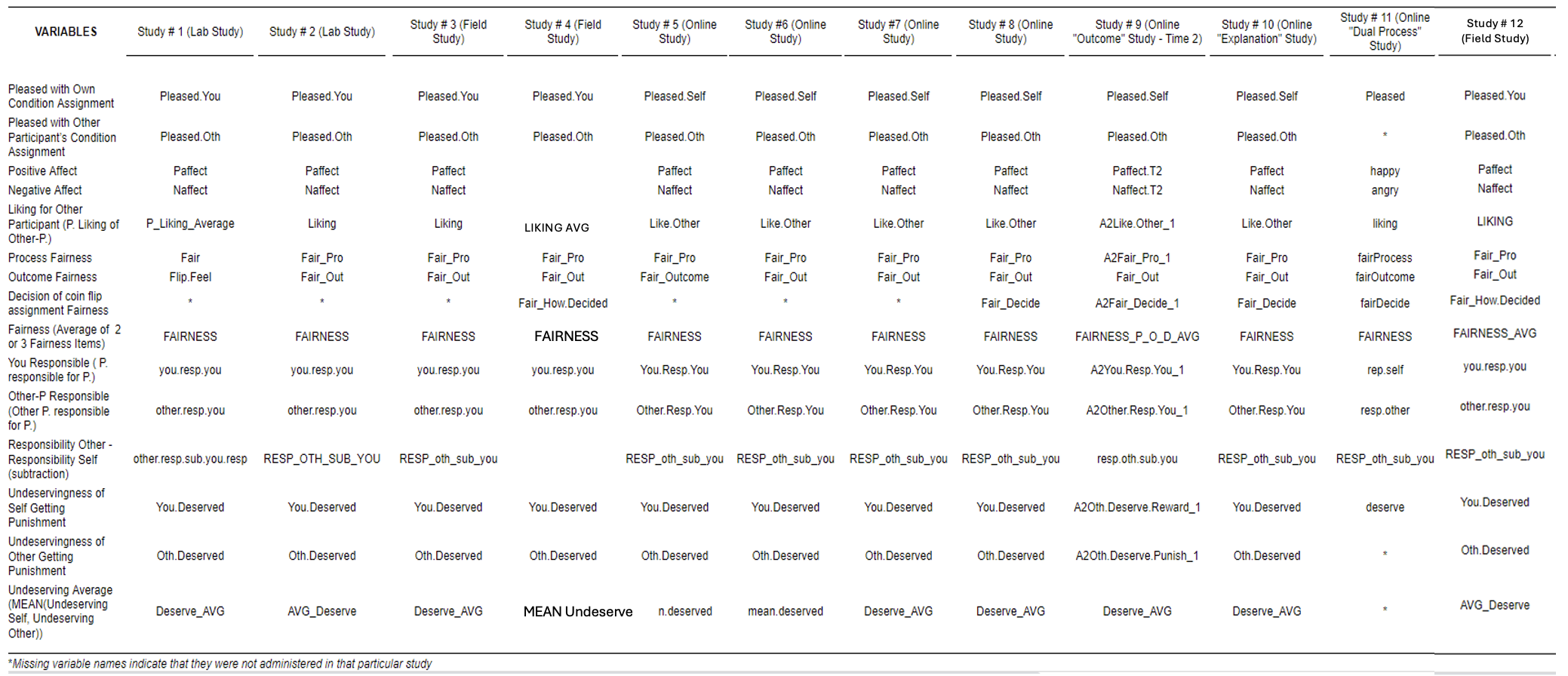
Study 8: <https://aspredicted.org/XZ9_JLR>

Study 9: <https://osf.io/4svmz/?view_only=1b2a94e9b2f541d290b83b38cc7e1f72>

Study 10: <https://aspredicted.org/MPF_BTF>  
Study 11: (1) <https://aspredicted.org/G29_4K3> (2) <https://aspredicted.org/C6Y_K8P>

Study 12: <https://aspredicted.org/ZJS_526>

# Outcome Variable Names Across the 11 Studies



Happy

\*

\*

# Data Inclusion

As stated in our pre-registrations, we included all participants in our analyses except those who failed manipulation checks of the condition they were in (i.e., those who were mistaken about who flipped the coin or who had received the bad outcome). The percentage dropped was generally low (e.g., none in Study 3 and five (2.3%) in Study 4), though it was somewhat higher in the online studies (ranging from 2.63% in Study 8 to 12.78% in Study 5). We dropped an additional eight participants; two from Study 1 and six from Study 2, due to experimental error (the participant accidentally won the coin toss). We also assessed participants’ level of suspicion, such as whether they believed that they would receive the negative outcome. Suspicion was quite low in the two lab and two field studies, but higher in the six online studies. As stated in our preregistrations we included suspicious people in all analyses, because dropping them had little impact on the results.

# Power Analyses

We ran an a priori power analysis using Study 1’s effect size of d = .57 for the basic illusion of unfairness effect using an independent samples *t*-test. The results suggested that a total of 100 participants was needed to achieve 80% power to detect the effect. We used this as a minimum benchmark for data collection in the lab and field studies (Studies 1-3).

Graphical user interface

Description automatically generated

Following data collection from Study 5 (the first online study), we ran an a priori power analysisusing Study 5’s effect size of d = .24 for the basic illusion of unfairness effect using an independent samples *t*-test. The results suggested that a total of 548 participants was needed to achieve 80% power to detect the effect. We used this as an approximate benchmark for data collection in the other online studies. For Studies 9, 10, and 11 we increased our sample size based on the maximum that was financially feasible with the goal of still detecting some potential interactions between conditions.

Graphical user interface

Description automatically generated

Here we report post hoc power analyses for Studies 1-10 based on *t*-tests run on the average fairness outcome variable. For Study 11 we report a post hoc power analysis based on a one-way ANOVA run on the average fairness outcome variable. All analyses were run using G\*Power.

**Study 1.**

Graphical user interface

Description automatically generated

**Study 2.**

Graphical user interface, application

Description automatically generated

**Study 3.**

Graphical user interface

Description automatically generated

**Study 4.**

**A screenshot of a test

Description automatically generated**

**Study 5.**

Graphical user interface

Description automatically generated

**Study 6.**

Graphical user interface

Description automatically generated

**Study 7.**

Graphical user interface

Description automatically generated

**Study 8.**

Graphical user interface

Description automatically generated with medium confidence

**Study 9.**

Graphical user interface, text

Description automatically generated

**Study 10.**

Graphical user interface

Description automatically generated

**Study 11.**

Graphical user interface, text, application

Description automatically generated

**Study 12.**

Graphical user interface

Description automatically generated

# *t*-test Results for the Illusion of Unfairness Across 9 Studies

In Table S2 and Table S3 we report a series of independent sample *t*-tests comparing the two conditions: P-Flipped and Other-P Flipped across 9 studies. Table S2 reports the basic results of the illusion of unfairness and Table S3 reports the affective consequences of the illusion of unfairness. We used these results to conduct the random effects meta-analysis reported in the manuscript (Table 1). Note that for nearly every outcome variable, the results yielded a significant Shapiro Wilks’ test, which suggests that these data violate normality assumptions. However, when comparing the results of a non-parametric test such as the Welch test and the standard *t*-test results we find very small differences (if any) between the Mann Whitney’s effect size (rank biserial correlations) and the Student *t*-test (Cohen’s d) when converting the two (see both effect sizes for each individual study in the *t*-test tables presented in this document). Furthermore, there is ample literature1,2,3,4,5 suggesting that using *t*-tests and ANOVAs to analyze non-normal data does not violate the validity of such significance tests, especially when the sample sizes and distribution shapes are similar across conditions, which is the case in our studies.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Table S2. Summary of *t*-test results for the basic illusion of unfairness effect (9 Studies). | | | | | | | | | | | | | | |  | |  |  | |  | | |  | | |
| Studies |  | Conditions |  | Fairness (Average) | | |  | | Other-P Responsible | | | |  | | You Responsible | | | | | | |  | | Undeservingness (Average) | | | | | |
|  |  |  |  |  | |  | |  |  |  | |  | |  | | |  | |  | | | |  | |  |  |  | |
|  |  | Mean (SD) |  | *d* | |  | | Mean (SD) |  | *d* | |  | | Mean (SD) | | |  | | *d* | | | |  | | Mean (SD) |  | *d* | |
|  |  |  |  |  |  |  | |  | |  |  |  | |  | |  | | |  | |  | | | |  | |  |  |  | |
| Study 1 (n=62) |  | Flip Self  (n= 32) |  | 6.39 (0.92) |  | 0.57\* | |  | | 1.26 (0.73) |  | 0.93\*\*\* | |  | | 2.34 (1.43) | | |  | | 0.27 | | | |  | | 3.84 (0.71) |  | 0.53\* | |
|  | Flip Other (n= 30) |  | 5.63 (1.66) |  |  | | 2.28 (1.39) |  |  | | 1.93 (1.66) | | |  | |  | | 4.38 (1.26) |  |
| Study 2 (n=150) |  | Flip Self (n= 74) |  | 4.45 (0.76) |  | 0.47\*\* | |  | | 1.32 (0.97) |  | 0.66\*\*\* | |  | | 2.32 (1.46) | | |  | | 1.09\*\*\* | | | |  | | 4.08 (0.71) |  | 0.21 | |
|  | Flip Other (n= 76 ) |  | 4.03 (0.96) |  |  | | 2.05 (1.22) |  |  | | 1.15 (0.48) | | |  | |  | | 4.24 (0.73) |  |
| Study 3 (n=75 ) |  | Flip Self (n= 38) |  | 4.18 (0.80) |  | 0.08 | |  | | 1.22 (0.67) |  | 0.68\*\* | |  | | 2.05 (1.22) | | |  | | 1.12\*\*\* | | | |  | | 3.46 (0.98) |  | 0.45 | |
|  | Flip Other (n= 37) |  | 4.10 (1.15) |  |  | | 2.03 (1.54) |  |  | | 1.05 (0.33) | | |  | |  | | 3.85 (0.76) |  |
| Study 5 (n=576) |  | Flip Self (n= 321) |  | 5.02 (1.67) |  | 0.24\*\* | |  | | 2.97 (1.42) |  | .33\*\*\* | |  | | 3.73 (1.43) | | |  | | 0.87\*\*\* | | | |  | | 4.25 (1.15) |  | 0.16 | |
|  | Flip Other (n= 255) |  | 4.59 (1.82) |  |  | | 3.45 (1.50) |  |  | | 2.57 (1.21) | | |  | |  | | 4.43 (1.10) |  |
| Study 6 (n=657) |  | Flip Self (n= 319) |  | 4.80 (1.83) |  | 0.30\*\*\* | |  | | 1.66 (1.17) |  | 0.58\*\*\* | |  | | 2.63 (1.45) | | |  | | 0.92\*\*\* | | | |  | | 4.20 (1.17) |  | 0.22\*\* | |
|  | Flip Other (n= 338) |  | 4.26 (1.80) |  |  | | 2.41 (1.42) |  |  | | 1.47 (1.01) | | |  | |  | | 4.46 (1.14) |  |
| Study 7 (n=346) |  | Flip Self (n = 172) |  | 5.04 (1.72) |  | 0.32\*\* | |  | | 1.76 (1.23) |  | 0.33 \*\* | |  | | 2.49 (1.42) | | |  | | 0.56\*\*\* | | | |  | | 4.18 (1.22) |  | 0.16 | |
|  | Flip Other (n= 174) |  | 4.47 (1.83) |  |  | | 2.19 (1.39) |  |  | | 1.74 (1.26) | | |  | |  | | 4.37 (1.07) |  |
| Study 8 (n=314) |  | Flip Self (n= 164) |  | 5.26 (1.48) |  | 0.45\*\*\* | |  | | 1.81 (1.32) |  | 0.37\*\* | |  | | 2.68 (1.45) | | |  | | 0.82\*\*\* | | | |  | | 4.07 (1.03) |  | 0.31\*\* | |
|  | Flip Other (n= 150) |  | 4.54 (1.70) |  |  | | 2.32 (1.45) |  |  | | 1.59 (1.22) | | |  | |  | | 4.40 (1.14) |  |
| Study 9 (n=1184) |  | Flip Self (n= 588) |  | 62.85 (24.57) |  | 0.35\*\*\* | |  | | 29.62 (27.08) |  | 0.50\*\*\* | |  | | 43.42 (29.61) | | |  | | 0.61\*\*\* | | | |  | | 53.75 (14.39) |  | 0.08 | |
|  | Flip Other (n= 596) |  | 53.83 (27.18) |  |  | | 44.55 (32.56) |  |  | | 25.47 (28.94) | | |  | |  | | 54.90 (15.46) |  |
| Study 10 (n=624) |  | Flip Self (n= 314) |  | 5.00 (1.62) |  | 0.23\*\* | |  | | 1.65 (1.16) |  | 0.41\*\*\* | |  | | 2.68 (1.42) | | |  | | 1.00\*\*\* | | | |  | | 4.16 (1.17) |  | 0.24\*\* | |
|  | Flip Other (n= 310) |  | 4.62 (1.73) |  |  | | 2.17 (1.39) |  |  | | 1.42 (1.06) | | |  | |  | | 4.42 (1.05) |  |
| Note *p* value significance levels are reported as \* on the *d* values. p<.05 = \*, p<.01 =\*\*, p<.001=\*\*\* | | | | | | | | | | | | | | |  | |  |  | |  | | |  | | |

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Table S3. Summary of *t*-test results for the downstream effects due to the illusion of unfairness effect (9 studies). | | | | | | | | | | | | | | | | | | | | | | | | | |
| Studies |  | Conditions |  | Pleased with own condition assignment | | |  | | Pleased with other P’s condition assignment | | | |  | Positive  Affect | | |  | | Negative  Affect | | |  | | | Liking for Other Participant | | | | | |
|  |  | Mean  (SD) |  | *d* | |  | | Mean  (SD) |  | *d* |  | Mean  (SD) |  | *d* |  | Mean  (SD) | |  | *d* | |  | Mean  (SD) | | |  | *d* | | |
| Study 1 (n=62) |  | Flip Self  (n= 32) |  | 2.78 (1.07) |  | 0.13 | |  | | 4.41 (0.88) |  | 0.29 |  | 2.64 (0.75) |  | 0.29 |  | 1.58 (0.70) | |  | 0.55\* | |  | -0.012 (0.73) | | |  | -0.18 | | | | |
|  | Flip Other (n= 30) |  | 2.63 (1.16) |  |  | | 4.07 (1.41) |  |  | 2.41 (0.82) |  |  | 2.03 (0.94) | |  |  | 0.103 (0.57) | | |  |
| Study 2 (n=150) |  | Flip Self (n= 74) |  | 2.89 (1.32) |  | 0.05 | |  | | 4.45 (1.35) |  | 0.38\* |  | 2.47 (0.90) |  | 0.06 |  | 1.62 (0.84) | |  | -0.04\* | |  | 4.42 (0.81) | | |  | -0.08 | | |
|  | Flip Other (n= 76) |  | 2.84 (1.28) |  |  | | 3.95 (1.26) |  |  | 2.42 (0.86) |  |  | 1.60 (0.82) | |  |  | 4.49 (0.84) | | |  |
| Study 3 (n=75) |  | Flip Self (n= 38) |  | 3.63 (0.85) |  | 0.56\* | |  | | 4.95 (1.11) |  | -0.05 |  | 2.99 (0.74) |  | 0.38 |  | 1.15 (0.32) | |  | 0.52\* | |  | 5.32 (1.01) | | |  | 0.18 | | |
|  | Flip Other (n= 37) |  | 3.16 (0.83) |  |  | | 5.00 (1.25) |  |  | 2.70 (0.79) |  |  | 1.49 (0.85) | |  |  | 5.14 (0.96) | | |  |
| Study 5 (n=576) |  | Flip Self (n= 321) |  | 2.63 (1.72) |  | 0.15 | |  | | 3.77 (1.57) |  | 0.22\*\* |  | 2.80 (1.06) |  | 0.19\* |  | 2.05 (0.95) | |  | -0.09 | |  | 4.82 (1.12) | | |  | 0.34\*\*\* | | |
|  | Flip Other (n= 255) |  | 2.39 (1.45) |  |  | | 3.43  (1.50) |  |  | 2.60 (0.98) |  |  | 1.97 (0.93) | |  |  | 4.44 (1.08) | | |  |
| Study 6 (n=657) |  | Flip Self (n= 319) |  | 2.27 (1.47) |  | 0.11 | |  | | 3.50 (1.60) |  | 0.16\* |  | 2.38 (1.02) |  | 0.02 |  | 2.16 (1.00) | |  | -0.02 | |  | 4.51 (1.11) | | |  | 0.14 | | |
|  | Flip Other (n= 338) |  | 2.13 (1.16) |  |  | | 3.26 (1.53) |  |  | 2.36 (0.93) |  |  | 2.14 (0.93) | |  |  | 4.36 (1.03) | | |  |
| Study 7 (n=346) |  | Flip Self (n= 172) |  | 2.46 (1.34) |  | 0.04 | |  | | 3.62 (1.43) |  | 0.12 |  | 2.60 (0.96) |  | 0.04 |  | 1.93 (0.91) | |  | -0.02 | |  | 4.45 (1.05) | | |  | 0.1 | | |
|  | Flip Other (n= 174) |  | 2.40 (1.34) |  |  | | 3.44 (1.58) |  |  | 2.56 (0.88) |  |  | 1.92 (0.92) | |  |  | 4.35 (1.11) | | |  |
| Study 8 (n=314) |  | Flip Self (n= 164) |  | 2.63 (1.60) |  | 0.16 | |  | | 3.63 (1.55) |  | 0.23\* |  | 2.70 (1.01) |  | 0.24\* |  | 1.97 (1.05) | |  | 0.05 | |  | 4.54 (1.08) | | |  | 0.2 | | |
|  | Flip Other (n= 150) |  | 2.40 (1.23) |  |  | | 3.28 (1.53) |  |  | 2.47 (0.90) |  |  | 2.00 (.99) | |  |  | 4.33 (1.07) | | |  |
| Study 9 (n=1184) |  | Flip Self (n= 588) |  | 20.25 (19.95) |  | 0.08 | |  | | 45.17 (24.24) |  | 0.35\*\*\* |  | 44.41 (22.07) |  | 0.02 |  | 31.38 (24.96) | |  | 0.08 | |  | 55.41 (15.52) | | |  | 0.20\*\*\* | | |
|  | Flip Other (n= 596) |  | 18.80 (18.34) |  |  | | 36.66 (24.75) |  |  | 43.94 (23.34) |  |  | 33.42 (26.62) | |  |  | 51.93 (18.74) | | |  |
| Study 10 (n=624) |  | Flip Self (n= 314) |  | 2.43 (1.31) |  | 0.12 | |  | | 3.62 (1.44) |  | 0.32\*\*\* |  | 2.52 (0.94) |  | 0.06 |  | 1.96 (0.91) | |  | 0.04 | |  | 4.48 (1.02) | | |  | 0.06 | | |
|  | Flip Other (n= 310) |  | 2.27 (1.33) |  |  | | 3.16 (1.43) |  |  | 2.46 (1.00) |  |  | 2.00 (0.96) | |  |  | 4.41 (1.09) | | |  |
| Note p value significance levels are reported as \* on the *d* values. p<.05 = \*, p<.01=\*\*, p<.001=\*\*\* . Study 4 is not included in this table because it only had win conditions and no lose conditions (see study 4 for results). | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  |  | |

# Meta-Analysis of the Illusion of Unfairness Results: Extended Methods and Results

**Variations across studies included in the meta-analysis.** This meta-analysis is comprised of all the studies we’ve run which participants competed against another participant and had the basic two conditions (P-Flipped and Other-P Flipped) necessary to test the illusion of unfairness. The fairness, liking and deservingness indices varied slightly between the studies in the following ways:

The fairness index of Studies 1, 2, 3, 5, 6 and 7 used the average of two fairness items measuring process and outcome fairness. Studies 8, 9, 10 used an additional fairness item which measuring how fair participants found the decision to be as to who got to flip the coin. In each study we created a fairness index by averaging the two or three fairness items that were included. The liking measures varied across studies in the following ways: Study 1 had participants complete 9 items about their liking for the other participant: How unlikeable/likeable, unpleasant/pleasant, unsociable/sociable, selfish/unselfish, inconsiderate/considerate, unfair/fair, incompetent/competent the other participant was, all on 9-point scales; how well they thought the other participant would fit in with their friends, and, after first meeting the other participant, how much most people would want to get to know him/her (the latter two on 5-point scales, *1 = not at all, 5 = very much*). These items were transformed into a *z*-score and averaged together. Study 2 had participants complete three items about their liking for the other participant: How unlikeable/likeable and how positive/negative is their impression of them, on 7-point scales, and how much they would want to work with this person on a class project on a 5-point scale (*1 = not at all, 5 = very much*). These items were also transformed into *z*-scores and averaged together. Studies 3-10 had participants fill out a single liking item: “If you were to meet the other participant, how much do you think you would like or dislike him/her?”, answered on a labeled 7-point scale (*1 = dislike a lot, 7 = like a lot*),

Study 9 had two time points at which participants filled out the dependent variables (once before the outcome of the coin flip was revealed and once afterwards). We used the dependent variables completed at the second time point in the meta-analysis, as they were the most similar to the other studies.

Study 12 was not included in the meta-analyses because the manipulation varied from the others. That is, instead of competing against another participant who was equally deserving of the reward, participants competed against the experimenter. If participants lost the flip they thus ended up helping the experimenter (by filling out a survey) for no compensation. If they won the flip they received appropriate compensation for their participation. Winning or losing was thus psychologically different than all the other studies. (At the end of the study, after being debriefed, all participants received the prize.)

Results from Study 10 demonstrated (using Bayesian equivalence tests) that there were no differences between explaining to participants that the coin flipper was chosen randomly via the count of odd or even numbers in their mTurk ID versus not giving an explanation (“we selected you/the other participant to flip the coin”). We therefore collapsed participants across those conditions and included them in the meta-analysis. The mTurk ID explanation was also used in Study 8 which was included in the meta-analysis.

Using JASP to conduct the random effects meta-analyses (Tables S3a and S3b) we ran a test of residual heterogeneity, which is calculated using Cochran's Q. When applying the random effects model, we used Restricted Maximum Likelihood (REML) to estimate the between-study variance (𝜏 2), which is factored into the study weights, to provide a more accurate reflection of both within-study error and unexplained heterogeneity across studies. The test of residual heterogeneity estimates whether the observed variability in effect sizes across studies is more than what would be expected based on sampling error alone. We find that the test of residual heterogeneity is not significant (except for outcome fairness and participants responsibility for their outcomes in Table S4a), which suggests that the random effects model tends to adequately account for the variability in effect sizes, and there is no significant remaining heterogeneity.”

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Table S4a. Meta-Analytic Effects of the Illusion of Unfairness Across 9 “Losing” Studies | | | | | |
| **Illusion of Unfairness Measures** | Overall  Meta-Analytic Effect Size | Omnibus Test of  Model Coefficients | | Test of Residual  Heterogeneity | |
|  | *d* [CI] | *Q* | *p* | *Q* | *p* |
| Fairness | -0.31 [-0.38, -0.25,] | 96.40 | <.001 | 6.7 | 0.57 |
| Procedural Fairness | -0.31 [-0.37, -0.24] | 92.99 | <.001 | 5.98 | 0.74 |
| Outcome Fairness | -0.25 [-0.35, -0.16] | 28.81 | <.001 | 14.49 | 0.07 |
| Decision Fairness | -0.34 [-0.43, -0.26] | 60.18 | <.001 | 0.93 | 0.63 |
| You Responsible | -0.82 [-0.96, -0.67] | 128.90 | <.001 | 30.76 | <.001 |
| Other-P Responsible | 0.46 [0.38, 0.54] | 131.76 | <.001 | 12.69 | 0.12 |
| Undeservingness | 0.19 [0.12, 0.26] | 29.00 | <.001 | 8.17 | 0.42 |
| **Downstream Affective Measures** | | | | | |
|  | | | | | |
| Pleased with Own Condition Assignment | -0.11 [-0.18, -0.05] | 12.91 | <.001 | 4.66 | 0.79 |
| Pleased with Other Participant’s Condition Assignment | -0.25 [-0.32, -0.18] | 44.90 | <.001 | 8.81 | 0.36 |
| Liking for Other Participant | -0.15 [0.24, -0.07] | 13.78 | <.001 | 11.65 | 0.17 |
| Positive Affect | -0.09 [0.16, -0.02] | 6.82 | 0.009 | 8.10 | 0.42 |
| Negative Affect | -0.03 [-0.09, 0.03] | 0.88 | 0.35 | 11.95 | 0.15 |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Table S4b. Meta-Analytic Effects of the Illusion of Unfairness Across 3 ‘Winning’ Studies | | | | | |
| **Illusion of Unfairness Measures** | Overall  Meta-Analytic Effect Size | Omnibus Test of  Model Coefficients | | Test of Residual  Heterogeneity | |
|  | *d* [CI] | *Q* | *p* | *Q* | *p* |
| Fairness | 0.21 [0.02, 0.40] | 4.88 | .03 | 0.63 | 0.73 |
| Procedural Fairness | 0.30 [0.11, 0.49] | 9.82 | 0.002 | 0.49 | 0.78 |
| Outcome Fairness | 0.12 [-0.12, 0.36] | 0.99 | 0.32 | 0.69 | 0.71 |
| You Responsible | -0.86 [-1.11, 0.61] | 46.22 | <.001 | 2.68 | 0.26 |
| Other-P Responsible | 0.73 [0.54, 0.92] | 54.61 | <.001 | 1.34 | 0.51 |
| Undeservingness | -0.17 [-0.35, 0.02] | 2.98 | 0.08 | 0.79 | 0.67 |
| **Downstream Affective Measures** | | | | | |
| Guilt | -0.26 [-0.45, 0.08] | 7.54 | 0.01 | 0.03 | 0.99 |
| Pleased with Own Condition Assignment | 0.05 [-0.14, 0.24] | 0.25 | 0.62 | 1.85 | 0.40 |
| Pleased with Other Participant’s Condition Assignment | 0.08 [-0.11, 0.26] | 0.62 | 0.43 | 0.28 | 0.87 |
| Liking for Other Participant | -0.02 [-0.21, 0.17] | 0.04 | 0.84 | 0.07 | 0.97 |
| Positive Affect | -0.06 [-0.25, 0.13] | 0.43 | 0.51 | 0.98 | 0.61 |
| Negative Affect  \* only in Study 2 and Study 3 | -0.05 [-0.30, 0.21] | 0.12 | 0.73 | 0.08 | 0.78 |

# Study 1

**Method**

**Overview**.Each experimental session consisted of a participant and a gender-matched confederate. The experimenter explained that they would each be performing a category learning task, with one member of the pair receiving a reward for each correct answer (reward condition), and the other receiving a punishment for each incorrect answer (punishment condition). Reward assignment (reward or punishment) was determined by coin flip, which was rigged such that the participant always lost and was assigned to the punishment condition. The independent variable was who flipped the coin: Participants were randomly assigned to conditions in which they flipped the coin, the confederate flipped the coin, or the experimenter flipped the coin.

**Participants.** Participants were 96 undergraduate students who participated in return for course credit. We dropped two from the analyses due to experimenter error (in these cases, the participants accidentally won the coin toss). Of the remaining 94 participants, 44% identified as female and 56% identified as male. Their reported races were 55% white, 22% Asian, 3% African American, 6% Hispanic, 12% Multi-Racial and 1% other. The mean age was 19.0 years (range = 18 to 22, *SD* = .97).

**Procedure.** Each session involved an experimenter, confederate, and participant. All groups were gender-matched; that is, the participant, confederate, and experimenter were either all female or all male. The confederate went to the waiting area a few minutes prior to the study start time and wore earphones in order to minimize any interaction with the participant. Once the confederate saw the participant arrive, they texted the experimenter to come greet them and bring them to the laboratory. After signing consent forms, the participant and confederate received written and verbal instructions explaining that they would each be performing a category learning task. On each trial they would see two drawings of fish, and be asked to judge whether they belonged to the same category. Through trial and error, they were told, they would gradually determine the correct classification rules. The purpose of the study (ostensibly) was to see whether people learn the rules faster when they are punishedfor wrong guesses or rewardedfor correct guesses.

The experimenter then took the participant and confederate to another room to further explain the different conditions for the category learning task. They were seated in chairs separated by a partition in order to minimize interactions. To further illustrate the reward condition, the experimenter showed the participant and confederate a basket of snacks (candy, nuts, fruit snacks) and said that in this condition, the participant would receive one snack for every correct answer. In the punishment condition, the experimenter explained, the participant would be required to put their hand in a bucket of ice water for 30 seconds after every incorrect answer. So that they would have a better idea of what this would entail, the experimenter removed a bucket of ice and water from a mini fridge and asked the participant and confederate to take turns submerging their hand for 15 seconds. The experimenter then emphasized that one participant would be in the positive reinforcement condition and that the other participant would be in the negative reinforcement condition, which would be determined by a coin flip.

***Coin flip conditions*.** A few moments earlier, while the participant and the confederate dried their hands with a paper towel, the experimenter discreetly looked at a notecard at the top of a stack of 3 randomly sorted cards, which assigned participants to condition. Both the experimenter and the confederate were thus unaware of conditions up to this point. In the confederate flipped condition (hereafter Other-P Flipped), the experimenter said to the confederate. “Hm . . . you can flip the coin,” followed by a shrug. He/she then asked the confederate whether he/she wanted heads or tails, and the confederate always answered heads. The experimenter reminded the confederate and participant that an outcome of heads would mean that the confederate would be in the reward condition and the participant would be in the punishment condition, whereas an outcome of tails would indicate the reverse outcome. The experimenter took out a quarter and demonstrated how to make a fist with the coin resting on the side of the index finger, such that it could be flipped by flicking the thumb. Because the coin had heads on both sides, the outcome of the flip always assigned the participant to the punishment condition. The procedure in the participant flipped condition (hereafter P-Flipped) was identical, except that the experimenter told the participant that he/she could flip the coin. If the participant called heads, the experimenter unobtrusively produced a coin with tails on both sides; if the participant called tails, the experimenter unobtrusively produced a coin with heads on both sides. This insured that the participant was always assigned to the punishment condition. In order to ensure that participants did not realize that they were flipping a double-sided coin, the experimenter would place the coin on the thumb of the participant. The procedure in the experimenter flipped (hereafter E-flipped) was identical, except that the experimenter said he/she would flip the coin. He/she said if the outcome was heads, the confederate would be in the reward condition and the participant in the punishment condition, whereas if the outcome was tails, the condition assignments would be the reverse. He/she then flipped the coin with heads on both sides, again ensuring that the participant was always assigned to the punishment condition. In sum, the participant was always assigned to the punishment condition via a coin flip; the only difference was whether the participant, the confederate, or the experimenter called heads or tails and flipped the coin.

***Dependent measures.*** After the coin flip the experimenter said that the participant and confederate would now fill out some baseline measures before doing the category learning task. They would do so, he/she said, on computers in separate rooms. After taking the confederate (and the snack basket) to another room, the experimenter returned and asked the participant to complete the questionnaire on a computer that was on an adjacent table. This questionnaire, which was programmed on Qualtrics, contained the dependent measures. Participants were first asked how pleased/displeased they were with the condition to which they had been assigned and the condition to which the other participant had been assigned, both on labeled 9-point scales (*1 = extremely displeased, 9 = extremely pleased*). As a measure of their affective state, participants then indicated the degree to which they were feeling calm, happy, nervous, distressed, irritated, angry, excited, relaxed, interested, and resentful, all on labeled 5-point scales (*1 = very slightly or not at all,* 5 *= extremely*). The emotion words were presented in random order.

Next, participants completed nine items about their liking for the other participant: How unlikeable/likeable, unpleasant/pleasant, unsociable/sociable, selfish/unselfish, inconsiderate/considerate, unfair/fair, incompetent/competent the other participant was, all on 9-point scales; how well they thought the other participant would fit in with their friends, and, after first meeting the other participant, how much most people would want to get to know him/her (the latter two on 5-point scales, *1* *= not at all, 5 = very much*). They then completed the same items according to how much they liked the experimenter.

Participants then answered five items about who was responsible for the condition to which they had been assigned: How responsible they were for the condition they received; how responsible the other participant was for the condition they received; how responsible the experimenter was for the condition they received; how responsible the other participant was for the condition he/she received, and how responsible the experimenter was for the condition the other participant received. Each of these items was answered on a labeled 5-point scale (*1 = not at all responsible, 5 = extremely responsible*).

Next participants answered two questions about fairness: How fair or unfair the procedure for deciding who was assigned to which condition was, and how fair the outcome of the coin flip felt to them. Both were answered on labeled 9-point scales (*1 = extremely unfair, 9 = extremely fair*). Lastly, participants indicated how much they felt they and the other person deserved the outcome they had received (the positive or negative reinforcement), each on a labeled 7-point scale (*1= extremely deserving,* *7= extremely undeserving*).

For another purpose, participants then completed the modern racism scale, a measure of the racial climate at their university, and a measure of their political orientation, followed by demographic questions.

After completing the dependent measures participants were told that the experiment was over and that they would not be doing the category learning task. They were fully debriefed with a funnel debriefing to gauge suspicion and offered a snack from the snack basket before departing.

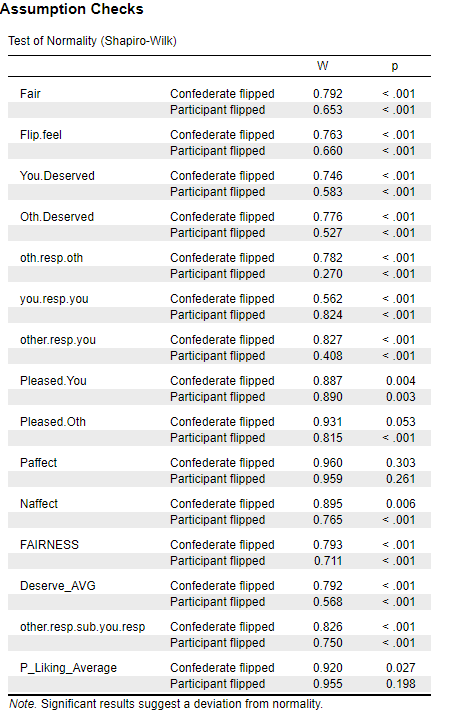
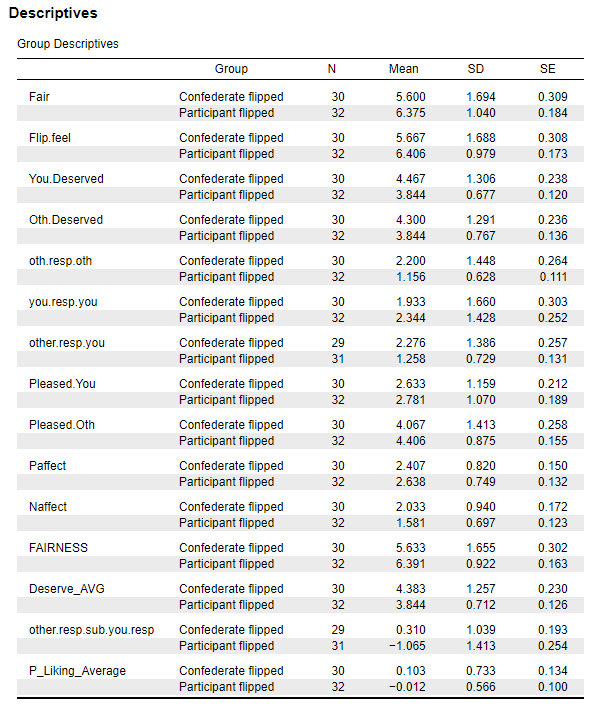
**Results**

For ease of presentation, we reduced the data by constructing scales. These scales were based on the results of this and the other studies, as well as our theoretical assumptions. In some of the individual scales the alphas were low; therefore, we report the results for all individual items in the supplementary materials. The scales were (1) their fairness ratings (the mean of how fair they thought the procedure and outcome was, alpha = .85); (2) how deserving they thought the outcomes were (the mean of their ratings of how deserving their outcome and the other participant’s outcome were, alpha = .91); (3) how responsible they thought they were for their outcome (single item); (4) how responsible they thought the other participant was (mean rating of Other-P’s responsibility for their own and the participant’s outcome, alpha = .91); (5) how responsible they thought the experimenter was (mean rating of the experimenter’s responsibility for their own and the participant’s outcome, alpha = .95); (6) a relative responsibility measure, which was the other participant’s rated responsibility minus the participant’s rated responsibility; (7) how pleased participants were the with the condition assignments (the mean of how pleased they were with their condition and the other participant’s condition, alpha = .18); (8) their negative affect (mean of their ratings of irritated, angry, distressed, nervous, and resentful, alpha = .85); (9) their positive affect (mean of their ratings of happy, calm, excited, relaxed, and interested, alpha = .74); (10) their liking for the other participant (mean of the nine standardized liking items, alpha = .83), and (11) their liking for the experimenter (mean of the nine standardized liking items, alpha = .83).

The descriptives and *t*-tests for each individual dependent measure are displayed in Table S5. For the purpose of the meta-analysis, we used the results of aggregated items, as described above.

## Table S5. Study 1 Descriptives and T-Tests

## Graphical user interface, application, table, Excel Description automatically generated



**Results of Experimenter Flipped Condition.** The question as to where the “action” lies within in the illusion of unfairness remains. That is, do people experience an increase in fairness when they flip the coin, a decrease in fairness when the other person flips the coin, or both? To address this question, we included a third condition in Study 1, in which the experimenter flipped the coin. In this condition the experimenter simply said, “I will be the one flipping the coin,” and arbitrarily assigned one of the participants heads and the other tails. We note that this was the first study we did and that it has the smallest sample size of the studies, and thus has limited power. Nonetheless it is interesting to see where the results of the experimenter flipped condition fall, relative to the other two conditions.

The results of Study 1 were largely consistent with the illusion of fairness hypothesis, on the measures of fairness, deservingness, and responsibility (at least how responsible the other participant was). As seen in Table S5 the effect sizes on these measures ranged from moderate (*d* = .47) to large (*d* = 1.08). There was less support for the hypothesis that who flipped the coin would have affective consequences; the contrast was significant only on the measure of negative affect. As seen in Table S6, the mean in the experimenter flipped condition fell in between the means in the P-Flipped and Other-P Flipped conditions on key variables. For example, when the experimenter flipped, participants’ fairness and undeservingness ratings, and their negative affect, were in between the means in the P-Flipped and Other-P Flipped conditions. Perhaps because the study had low power, the means in the experimenter flip condition were not significantly different from the means in either the P-Flipped or Other-P Flipped condition (with the exception of ratings of how responsible the other participant was for the outcome, see Table S6). Nonetheless, the results suggest that the “action” of the illusion of unfairness resides in both of the experimental conditions: Flipping the coin oneself seems fairer, and seeing the other person flip the coin seems less fair.

## Table S6. Experimenter Flip vs P-Flipped vs Other-P Flipped Contrasts

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Dependent Measure** | **Participant Flip**  ***n***  ***SD***  ***M*** | **Exper. Flip**  ***n***  ***SD***  ***M*** | **Other-P Flipped**  ***n***  ***SD***  ***M*** | **Omnibus**  **Test**  ***F***  ***df***  ***p*** | **Contrast**  **(1 0 -1)**  ***t***  ***p***  ***d*** |
| Fairness Mean | 32  .92  6.39 | 32  1.19  5.98 | 30  1.66  5.63 | 2.70  91  .072 | 2.32  .023  .49 |
| Deservingness  (Hi = Undeserving) | 32  .71  3.84 | 32  1.05  4.19 | 30  1.26  4.38 | 2.22  91  .115 | -2.07  .041  .43 |
| P Responsible | 32  1.43  3.34 | 32  1.12  2.63 | 30  1.66  2.93 | 2.07  91  .13 | 1.14  .257  .239 |
| Other P Responsible | 32  .65  1.20 | 32  .83  1.42 | 30  1.38  2.27 | 9.86  91  <.001 | -4.22  <.001  . 884 |
| Experimenter’s Responsibility | 32  1.12  1.81 | 32  1.33  2.03 | 30  1.10  1.75 | .484  91  .618 | .733  ..465  .154 |
| Other-P Responsible – P Responsible | 31  1.38  -1.14 | 32  .80  -.20 | 30  1.12  .33 | 13.70  91  <.001 | 5.15  <.001  1.080 |
| Pleased with Condition Assignments | 32  .71  3.59 | 32  .91  3.28 | 30  .95  3.35 | 1.16  91  .32 | 1.12  .268  .234 |
| Negative Affect | 32  .70  1.58 | 32  .71  1.84 | 30  .94  2.03 | 2.59  91  .080 | 2.26  .026  .47 |
| Positive Affect | 32  .75  2.64 | 32  .75  2.33 | 30  .82  2.41 | 1.40  91  .25 | 1.17  .24  .25 |
| Liking for Other P | 32  .57  -.012 | 32  .66  -.044 | 30  .73  .103 | .43  91  .65 | -.70  .489  .15 |
| Liking for Experimenter | 32  .57  .059 | 32  .61  -.069 | 30  .81  -.049 | .34  91  .72 | .64  .526  .13 |

# Study 2

**Method**

**Overview**.Each experimental session consisted of two participants. The experimenter explained that they would each be performing a category learning task, with one member of the pair receiving a reward for each correct answer (reward condition), and the other receiving a punishment for each incorrect answer (punishment condition). Condition assignment (reward or punishment) was determined by coin flip. The independent variable was who flipped the coin: Participants were randomly assigned to conditions in which they flipped the coin. This study employed a 2 x 2 design: flip/no-flip vs win/lose conditions. We report the demographics and results for the win/lose conditions separately below.

**Participants: Lose conditions.** Participants were 156 undergraduate students who participated in return for course credit. We dropped six from the analyses due to participant errors in misinterpreting the outcome (win/lose) of the coin flip. Of the remaining 150 participants, 77% identified as female and 72% identified as male, and one participant identified as transgender. Their reported races were 72% white, 31% Asian, 2% Hispanic, 1% Pacific Islander, 0.6% African American, and 1% other (participants could select to report multiple races). The mean age was 19.0 years (range = 17 to 28, *SD* = 1.25).

**Participants: Win conditions.** Participants were 154 undergraduate students who participated in return for course credit. Across the participants, 51% identified as female and 49% identified as male, and one participant identified as transgender. Their reported races were 62% white, 32% Asian, 8% Hispanic, <1% Pacific Islande, <1% Native American, 7% African American, and 3% other (participants could select to report multiple races).. The mean age was 19 years (range = 17 to 51, *SD* = 2.84).

**Procedure.** The procedure was identical to that of Study 1 except for these changes: Two actual participants were run in each session, instead of employing a confederate. The coin toss used a real coin, and in this manner one participant was randomly assigned to the punishment condition and the other the reward condition. Participants were told that if they received the punishment condition, they would have to put their hands in a bucket of ice water for 15 seconds for each incorrect trial (compared to 30 seconds in Study 1). To minimize experimenter bias, the experimenter informed participants that he/she could not be made aware of the coin flip’s outcome, so the coin flipper had to let the coin fall onto the ground and both participants would look at the result without interacting with each other or touching the coin. Prior to the coin flip procedure, the experimenter would grab the snack basket and leave the room while one participant flipped the coin. Before entering the room again, the experimenter would ask participants if the coin flip procedure was complete and would then pick up the coin without looking at the outcome. In this manner the experimenter was unaware of participants’ condition.

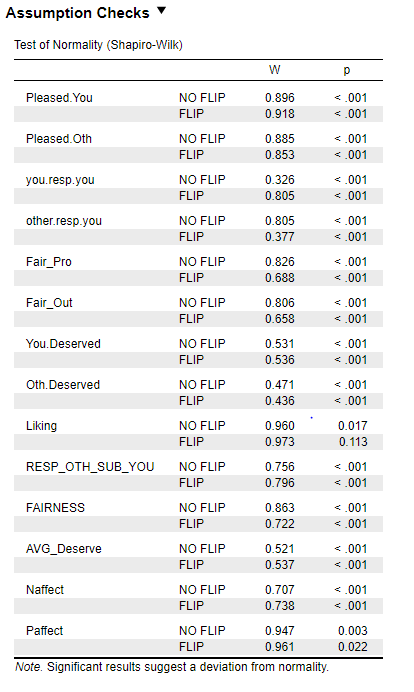
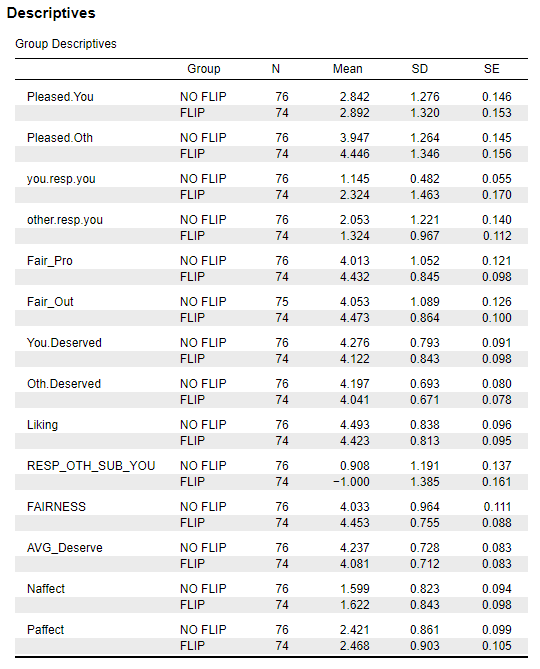
***Dependent measures*.** The dependent measures were identical to those of Study 1 except for: Liking for the other participant, guilt, and liking of the other participant was measured using three items: “If you were to spend more time with the other participant, how much do you think you would like or dislike him/her? (*1=Dislike A Lot, 7= Like A Lot*)”, “How would you rate your initial impressions of the other participant, based on your interactions with him/her today? (*1= Extremely Negative, 7= Extremely Positive*)”, “Compared to other students, how much would you like to work with this person on a class project? (*1* *= not at all, 5 = very much*)”. Since the first two items listed above were measured on 7-point scales we standardized the three liking items before averaging them together.

We computed the same two item fairness scales as reported in the first study, which had an alpha reliability rating of .822, two items for deservingness (.696), 5 items for negative affect (.802), 5 items for positive affect (.754) and three items for liking (.829).

**Results: Lose Condition**

The results of a series of independent samples *t*-tests comparing the P-Flipped vs Other-P Flipped conditions are reported below in Table S7.

## Graphical user interface, application, table, Excel Description automatically generatedTable S7. *t*-test results Study 2, Lose Condition



**Results: Win Condition**

The results of a series of independent samples *t*-tests comparing the P-Flipped vs Other-P Flipped conditions are reported below in Table S8.

## Table S8. *t*-test results Study 2, Win Condition

A screenshot of a computer

Description automatically generated

A table with numbers and text

Description automatically generatedA table with numbers and text

Description automatically generated

# Study 3

**Method**

**Overview**.This study was a field study run on the campus at a University. Its aim was to replicate our lab study in a different setting with different incentives. Because participants in the lab study ultimately participated in the study with the purpose of receiving guaranteed course credit, we wanted to test the illusion of fairness in a context where there would be no reward at all for the participant who lost the coin flip. This study employed a 2 x 2 design: flip/no-flip vs win/lose conditions. We report the demographics and results for the win/lose conditions together below.

**Participants.** 150 participants (75 pairs) completed the study. Of the 150 participants, 76 participants lost the coin flip and 74 participants won the coin flip. Of those participants, 72 identified as male and 70 identified as female, 2 reported identifying as “gender fluid”. Participants’ age ranged between 17-44 years old with a mean of 19.5 years. Participants self-identified their race as follows: 86 White, 29 African American, 29 Asian, 10 Hispanic, 0 American Indian/Alaskan Native, 1 Pacific Islander, 1 Russian and 1 Mixed. All participants correctly reported having won/lost as well as having flipped the coin or not. We included all participants in our analyses because when we dropped: 3 participants who reported having met the other participant, but not knowing them very well, 8 participants who reported recognizing the person but never having met, 10 participants who offered their prize to the other participant, and 10 participants who reported having completed one of the lab studies regarding the illusion of unfairness, we found no major differences compared to including them in the final dataset. We therefore decided to keep all participants.

**Procedure.** Participants were recruited from a street adjacent to the university as well as in front of one of the dining commons. The experimenter said to passersby: “Would you like the chance to win a cold drink for completing a brief survey?”. If they were interested in participating, we told them: “We need two people who don’t know each other. Only one of you will get the drink and we’ll decide who gets the prize with a coin flip. Do you want to participate?” This was followed by instructions to wait until we found another participant whom they did not know. Once both strangers were ready, we sat them down facing each other and read to them the following instructions: “Before we begin, we ask that you do not interact or speak with each other until after you are finished filling out the survey. Here’s how it will work. First, we’ll flip a coin to see who gets the prize. Then we will ask each of you to complete our brief survey on these tablets, which will take about 5 minutes.”

Following the instructions participants were told, “OK, let’s flip the coin. As you can see, it’s a fair coin, with heads on one side and tails on the other. Whoever wins will get the prize at the end of the study.” We then gave them instructions as to how to flip a coin correctly: “Whoever flips the coin will have to extend their arm in front of them and make this shape with their hand. They will place this coin between their thumb and index finger, to make sure they complete a standard coin flip. They will then proceed to flip it by simply flicking the coin with their thumb while keeping their wrist stable and letting the coin fall on the floor.” The experimenter, unbeknownst to the participants, assigned heads to the participant sitting on their right and tails to the participant on their left. The experimenter then demonstrated a coin flip and used the outcome to decide who would get to flip the coin. The experimenter looked at the coin in their hand, paused for half a second and without smiling, looked between participants then moved their gaze to either the participant on their left or right. The experimenter would then point with their clipboard and say: “you can flip the coin.” The experimenter then proceeded to say: “Do you want heads or tails? Okay, we know you'll be either happy or disappointed if you win or lose, but please try to remain silent.” They then summarized the procedure and potential outcomes.

Following the participant’s coin flip, the experimenter said: “OK, you [pointing to the winner] will get the prize. Go ahead and choose one of the drinks or snacks. Now I have a brief survey for you each to fill out, it won’t take long.” Participants then proceed to fill out the key dependent variables.

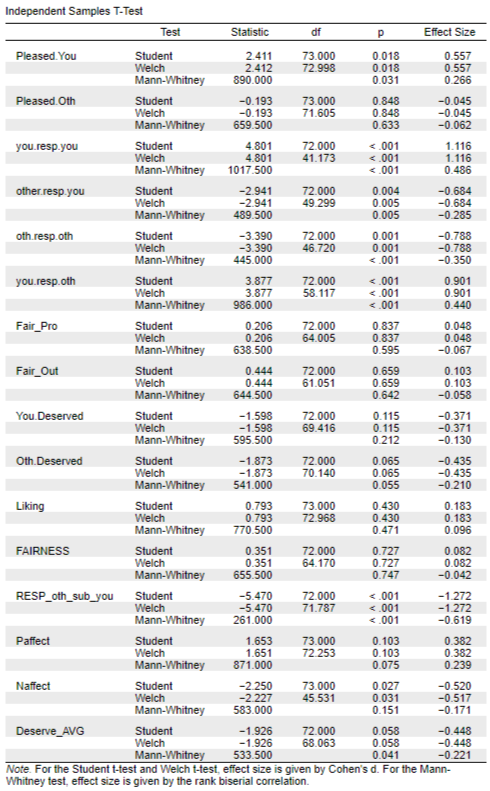
**Results**

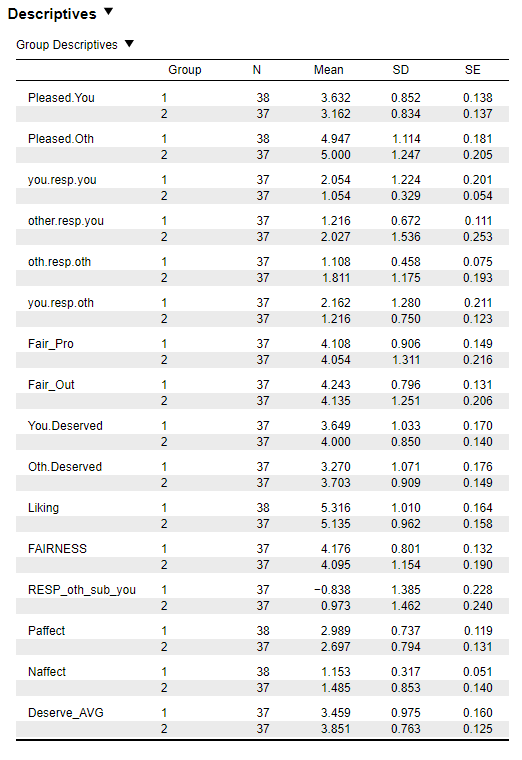
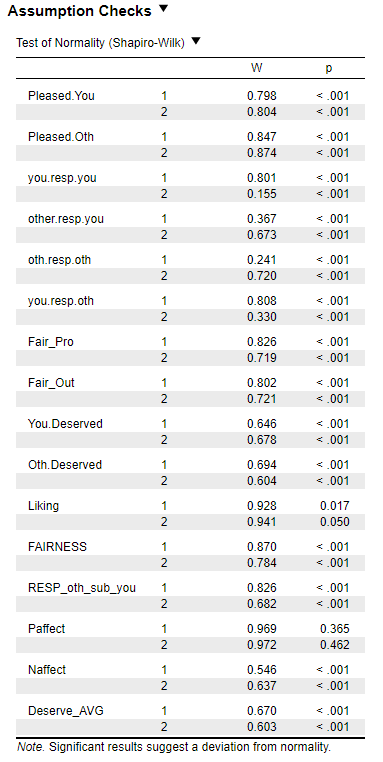
The following dependent variables were averaged: 2 Fairness items (process and outcome) ratings (.850), 5 items negative affect (.782), and 5 items positive affect (.690), two items for deservingness (.738) Other dependent measures (e.g., responsibility and liking) were single items.

**Lose Condition**

The descriptive results and the independent samples *t*-tests comparing the P-Flipped vs Other-P Flipped conditions are reported below in Table S9.

## Table S9. Descriptive and *t*-test Results of Study 3





**Win Condition**

A table with numbers and text

Description automatically generated

**A table of text with black text

Description automatically generated with medium confidenceA table with numbers and text

Description automatically generated**

# Study 4

**Method**

**Overview**. This field study was run in a small park on the campus at a university. Its aim was to provide a further test of how participants experienced winning the coin flip.

**Participants**. Participants were passersby recruited from a street adjacent to a university park. They completed the study at a table in the park. Of the 217 participants, 85 identified as male, 130 identified as female, two identified as other. Participants’ age ranged from 17 - 63 years old with a mean of 20.68 years. Participants self-identified their race as follows: 129 White, 16 African American, 69 Asian, 11 Hispanic, 1 Native American, 1, Mixed, 3 other. We removed 5 participants from the analyses: three were given the wrong coin to flip, one participated twice, and one noticed that the coin was double-sided.

**Procedure.** The procedure was the same as in the first field study except that the participant was competing for the prize with a confederate who posed as the second participant. Further, the actual participant always won the coin toss, which was accomplished by using a two-sided coin. The experimenter first recruited a passerby from the street to complete a survey, and while the participant waited at the bench, the experimenter recruited another participant (the confederate) from the street (this ensured that participants weren’t suspicious of a confederate simply waiting for the study to start). The experimenter stated “OK, let’s flip the coin. While saying this, they checked the time on their phone in their hand, if the minutes showed an odd number the confederate would flip, and if the minutes were even, the participant would flip. This ensured that the confederate and experimenter were unaware of the experimental conditions up until this point. When the confederate flipped the coin, they would always choose heads, and the experimenter would place the double-tailed coin on their hand (to ensure the participant won). When the participant flipped the coin, the experimenter would always place a double-sided coin that matched their selection of heads or tails (to ensure that they won). Participants then completed the dependent variables on a tablet.

**Results**

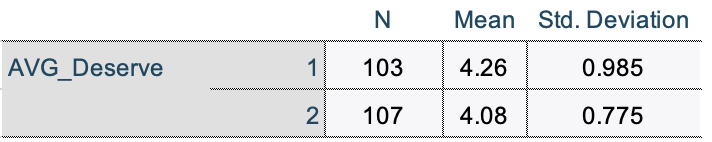
The following dependent variables were averaged: Three fairness items (process, outcome, and decision; alpha = .79), two items for liking (.67), and two items for deservingness (.80) Other dependent measures (e.g., negative affect and responsibility) were single items.

The descriptive results and independent samples t-tests comparing the P-Flipped vs Other-P Flipped win conditions are reported below in Table S10. Note that P-flip is labeled as 1 and Other-P-flip is labeled as 2.

## Table S10. Descriptives and *t*-test results Study 4 (winning outcome)

*A screenshot of a computer

Description automatically generatedNote:* AVG\_Deserve was also computed for the Win meta-analysis (*d*= 0.2, SE=0.14).



# 

# Studies 5 and 6

**Method**

**Overview**. Participants on mTurk were told that another participant would be taking part at the same time. The instructions were identical to Study 1, namely that they and the other participant would be performing a category learning task, that one would be assigned to a reward condition and the other to a punishment condition, and that the outcome each received would be determined by a coin flip. The details of how the other participant was simulated varied across studies, as described below. In Studies 5 and 6, participants were randomly assigned to a P-Flipped condition, in which they flipped a virtual coin, or to an Other-P Flipped condition, in which the other participant (ostensibly) flipped the virtual coin. As in Study 1, the real participant lost the coin flip in both conditions, and was thus assigned to the punishment condition. The dependent measures were the same as in Study 1. We note that both studies included an additional manipulation, namely whether the ostensible other participant was White or Black, to see whether their race influenced the magnitude of the illusion of unfairness. In Studies 5, participants exchanged demographic information and learned the race of the other participant; in Study 6, participants learned that the first name of the other participant was Greg or Jamal, which are stereotypically White or Black names. As it happened, there were very few significant main effects of race of the other participant or Race x Who Flipped interactions, no more than would be expected by chance. We thus averaged across race of the other participant in these studies.

**Participants.** Participants were recruited from Amazon Turk on Sunday, August 18, 2019; Saturday, August 31, 2019; and Sunday, September 1, 2019 for Study 5, and on Saturday, Dec. 21, 2019; Sunday, Dec. 22, 2019; Monday, Dec. 23, 2019; Sunday, Dec. 29, 2019; Monday, Dec. 30, 2019; and Saturday, Jan. 4, 2020. We deliberately ran the studies on weekends in order to lower the likelihood that participants completed the study while at work. Participants were told that they would receive $2 for a study taking approximately 20-30 minutes (in Study 5) or $1.50 for a study taking approximately 15-25 minutes (in Study 6). As described below, all participants were given a bonus of $1 at the end of the study, thus, participants who completed the study were paid a total of $3 (Study 5) or $2.50 (Study 6).

Of the 1,120 participants who began Study 5, 399 dropped out or were dropped before being assigned to condition, for one of these reasons: They were doing the study on a smartphone (we required that they be using a laptop); they failed a bot detection question; they reported that they didn’t have time or weren’t alone; they preferred not to share their screen with another participant; or for unknown reasons. Of the 721 who completed the dependent measures, 380 identified as male, 339 identified as female, and 2 identified as other. Participants indicated that they were in the following age ranges: 18-25 (76), 26-39 (391), 40-59 (196), and 60 or older (58). Participants self-identified their race as follows: 527 White, 111 Black, 51 Asian, 16 Latino/a, 1 American Indian/Alaskan Native, and 14 Multiracial.

In Study 6, of the 998 participants began the study, 295 dropped out before being assigned to condition for the same reasons as in Study 5. Of the remaining 703 participants, 344 identified as male, 355 identified as female, and 4 identified as other. Participants indicated that they were in the following age ranges: 18-25 (84), 26-39 (378), 40-59 (189), and 60 or older (52). Participants self-identified their race as follows: 509 White, 75 Black, 41 Asian, 41 Latino/a, 6 American Indian/Alaskan Native, and 21 Multiracial.

**Procedure.** After consenting to participate, and answering a question designed to identify bots, participants responded to some preliminary questions, such as whether they had 20-30 minutes to commit to the study, whether they were alone, and whether they had turned off all devices that could be distracting. They then learned that they would be paired with another mTurk participant. In Study 5 they were told that they and the other participant would be asked to share what was shown on their computer screens with each other, and were given the option to opt out of the study if they preferred not to share their screens (72 participants did so). There was no screen sharing in Study 6.

Participants then learned that they would be performing a category learning task, receiving instructions identical to those delivered in Study 1. As in that study, they learned that either they or the other participant would be assigned to a reward condition, whereas the other would be assigned to a punishment condition. The participant assigned to the reward condition, they read, would receive $.10 after every correct answer up to a bonus of $1.00. The participant assigned to the punishment condition, they read, would, after every wrong answer, be asked to read a paragraph from a textbook and summarize that paragraph in four sentences. (This activity was rated by pilot mTurk participants as particularly undesirable.) Participants answered a series of comprehension check questions to make sure they understood the instructions; if they answered incorrectly, the instructions were repeated.

Participants then learned that a “virtual coin flip” would determine which participant would be assigned to the reward condition and which would be assigned to the punishment condition. “The way it works,” they read, “is that you or the other participant will select Heads or Tails and then click on a virtual coin, which will start spinning.” If the person who selected heads or tails won the toss, that person was assigned to the reward condition; if they lost the toss, they were assigned to the punishment condition. The way in which the other participant was portrayed differed slightly across the two studies:

***Study 5.*** Participants were told that “we alternate which person gets to select Heads/Tails and flip the coin,” and that that person’s computer screen would be shared with the other participant, so that he/she could observe the coin flip. Participants were asked to wait while a connection was established. “This may take a while,” they were told, “depending on the availability of other mTurk participants.” A throbber appeared on the screen, to (supposedly) indicate that the program was waiting for another participant to connect. After 20 seconds, the screen read, “Establishing connection with mTurker { A3IZSXSSGW8OFN } ,” followed by another throbber that remained on the screen for 8 seconds. Participants were asked to indicate their gender, age range, and race, which would supposedly be sent to the other participant, who would respond in kind. Each participant would then be asked to verify what they learned about the other participant, which would supposedly serve to eliminate bots. After entering their information, participants learned that the other participant was a male who was 26 – 39 years old. For another purpose, we manipulated the race of the other participant; he was described as “White/Caucasian” or as “Black/African American” (randomly assigned). On the next screen, participants were then asked to verify that information by indicating the gender, age, and race of the other participant. As it happened, there were no significant effects of the race of the other participant on any dependent measure, with very few exceptions. Therefore, we collapsed across this variable in the analyses reported here.

Participants then learned either that they or the other participant (randomly assigned) “will select Heads/Tails and do the virtual coin flip.” In the participant flip condition (hereafter called P-Flipped), they were given instructions to enlarge their screen so that the other participant could see it clearly, saw a screen that supposed launched screen sharing, selected Heads or Tails, and then pressed a button that started a picture of a U. S. quarter spinning. All participants lost the coin toss, assigning them to the punishment condition. That is, if they selected Heads the coin landed on Tails, and if they selected Tails the coin landed on Heads. In the other participant flip condition (hereafter called Other-P Flipped), participants were given instructions to enlarge their screens so that they would get a good view of the other participant’s screen, saw a screen that supposed launched screen sharing, and then saw the other participant selecting tails and pressing a button that started a picture of a U. S. quarter spinning (they were actually watching a video). The coin always landed on tails, such that once again, the participant was assigned to the punishment condition. In other words, in both the P-Flipped and Other-P Flipped conditions, participants were assigned to the punishment condition whereas the other participant was assigned to the reward condition. The only difference was who got to call heads/tails and spin the virtual coin.

All participants then saw a screen that supposedly disconnected the screen sharing and answered the following manipulation check question: Who got to choose heads/tails and flip the coin; whether the coin landed on heads or tails, and whether they would be in the punishment or reward condition.

***Study 6.*** The procedure in Study 6 was identical except for these differences: After learning that another participant had (supposedly) joined the study, participants were then told that a botcheck would be presented in which they would answer questions about two paragraphs from Wikipedia. The first paragraph was about tooth brushing, and participants typed an answer to the question, “What hardens on your teeth if you don’t brush every 24 hours?” The second paragraph was about affirmative action, and participants typed an answer to the question, “Who is given a preference in selection processes, under affirmative action?” After learning that both they and the other participant has passed the bot check, participants read that “we like to personalize things a bit by having each of you enter a little information that will be shared with each other, including your first name, your gender, and your age range.” They were asked to enter their first name, gender, and age range (18-25, 26-39, 40-59, or 60 and above), after which they saw what the other participant has supposedly entered. Participants saw that the other person was a male aged 26-39. Once again, for another purpose, we manipulated the race of the other participant, this time by varying his name. Half of the participants learned that his name was Jamal, a stereotypical Black name, whereas half learned that his name was Greg, a stereotypical White name. Participants were asked to verify what they learned by entering the other person’s name, gender, and age range on the next page. And, once again, the race of the other participant had no significant effect on any dependent measure, with very few exceptions. Therefore, we again collapsed across this variable in the analyses reported here.

Participants were randomly assigned to the P-Flipped or Other-P Flipped condition in the same manner as in Study 5, except that there was much less ostensible screen sharing. In the P-Flipped condition, participants flipped the coin and lost. In the Other-P Flipped condition, participants were told that the other participant had chosen tails, then saw an animation that supposedly depicted the other participant pressing a button that started a picture of a U. S. quarter spinning and then landing on tails.

***Dependent measures****.* After a filler question asking participants how pleased or displeased they were about having to share their screen with the other participant, participants answered the same dependent measures as assessed in Study 1, with these exceptions: Because there was not a condition in which an experimenter flipped a coin, all questions about the role of the experimenter were dropped. For length reasons we reduced the question about liking for the other participant to a single item, “If you were to meet the other participant, how much do you think you would like or dislike him/her?,” answered on a labeled 7-point scale (*1 = dislike a lot, 7 = like a lot*), and reduced the responsibility questions to two of the items asked in Study 5: how responsible participants felt they were for the outcome they received, and how responsible they felt the other participant was for the outcome they received. Participants then completed the Symbolic Racism Scale (for another purpose), and then, in random order, a system justification scale and locus of control scale.

Participants then answered six open-ended questions asking for their general impressions of the study and assessing their suspicions (e.g., “What did you think about how it was decided that either you or the other participant would get to choose Heads or Tails and flip the virtual coin?”). They then completed rating scales assessing their suspicions more directly (e.g., “How much did you think there would be another mTurk participant taking part in this study at the same time?”, answered on a 5-point scale where *1 = not at all* and *5 = completely*). After providing some further demographic information, participants learned that the study was over and that they would not be doing the category learning task. They read a detailed explanation of the study, including the fact that there was not actually another participant and that there was not actually any screen sharing. This was followed by a post-debrief consent form in which participants were given the option of emailing the researchers and asking that their data be destroyed (none did so). Finally, all participants learned that they would receive a $1 bonus and given the opportunity to provide any further comments.

**Results**

There was a fair amount of reported suspicion about various aspects of the procedure in both studies. When asked whether they thought the virtual coin spin was random, 16% of participants in Study 5 and 23% of participants in Study 6 responded “not at all.” When asked whether they thought there was another participant taking part at the same time, 11% of participants in Study 5 and 20% of participants in Study 6 responded “not at all.” As it happened, however, when we dropped suspicious participants from the analyses, the results changed little. We suspect this is because at least some participants were not very suspicious when completing the dependent measures but became more so after being probed with specific questions. We thus opted to retain participants in the analysis regardless of their suspicion ratings.

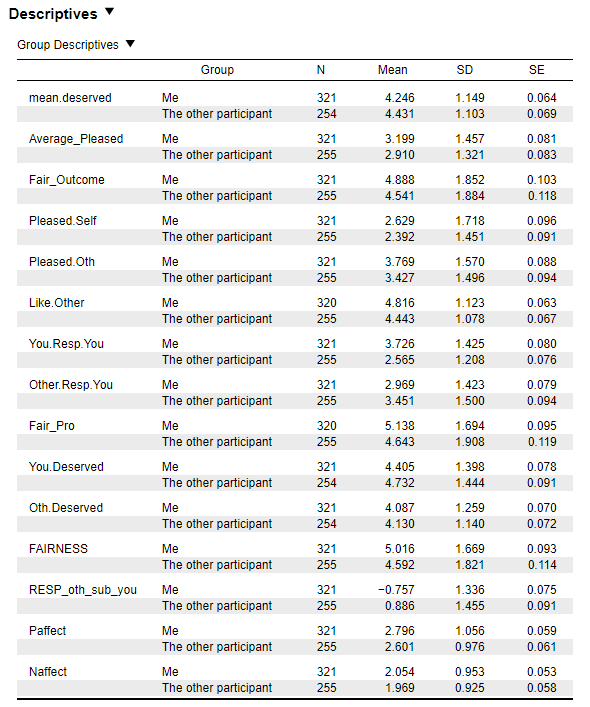
Consistent with our preregistration, we did drop participants who failed either one or both of two key manipulation checks: Recalling who got to flip the coin and recalling whether they would be in the punishment or reward condition. In Study 5, 145 participants answered one or more of these questions incorrectly. More did so in the Other-P Flipped condition (100) than in the P-Flipped condition (45), a difference that was significant, *X2*(1, *N* = 721) = 28.26, *p* < .001. In Study 6, only 46 participants answered one or both questions incorrectly. Again, more did in the Other-P Flipped condition than the P-Flipped condition, 31 vs. 15. *X2* (1, *N* = 703) = 6.10, *p* = .014, the difference was less pronounced than it was in Study 5.

We computed the same scales as reported in previous studies which had these alphas for Studies 5 and 6, respectively: Fairness ratings (.887, .893), deservingness (.637, .713), how pleased participants were (.739, .697), negative affect (.871, .869), and positive affect (.878, .876). Other dependent measures (e.g., liking for the other participant) were single items.

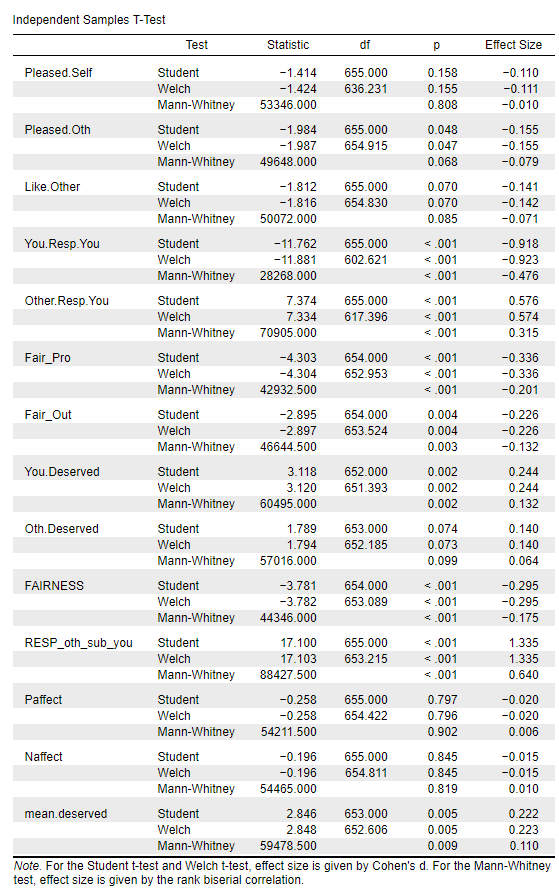
As seen in Tables S9 and S10, the results were as expected on the majority of the dependent variables.

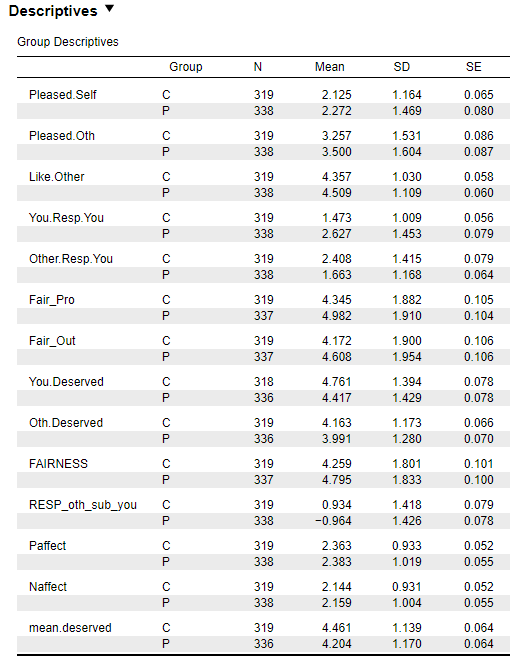
## Table S11. Descriptive and t-test Results of Study 5





## Table S12. Descriptives and *t*-test Results of Study 6





# Study 7

**Method**

**Participants.** Participants were recruited from Cloud Research on Sunday, April 5, 2020. we limited participation to those with an approval rating of at least 90% and hits completed between 100 - 50,000. Of the 995 participants who began the study, 278 dropped out or were dropped before being assigned to condition, for one of these reasons: They were doing the study on a smartphone (we required that they be using a laptop); they failed a bot detection question; they reported that they didn’t have time or weren’t alone; or for unknown reasons. Of the 717 who completed the dependent measures, 374 identified as female, 339 identified as male, and 4 identified as other. Participants indicated that they were in the following age ranges: 18-25 (95), 26-39 (334), 40-59 (207), and 60 or older (81). Participants self-identified their race as follows: 538 White, 65 Black, 55 Asian, 31 Latino/a, 3 American Indian/Alaskan Native, and 12 Multiracial. As discussed below, we dropped 50 participants from the analyses who failed one or more manipulation checks. Of the remaining 667 participants, 365 identified as female, 308 identified as male, and 4 selected other. These participants indicated that they were in the following age ranges: 18-25 (92), 26-39 (304), 40-59 (202), and 60 or older (79), and self-identified their race as follows: 515 White, 57 Black, 52 Asian, 30 Latino/a, 3 American Indian/Alaskan Native, and 10 Multiracial.

**Procedure.** The procedure was identical to Study 6 except for these changes: (1) We did not manipulate the race of the other participants. Participants were asked to enter their initials followed by a single digit of their choice; they learned that the other participant had entered ND8, was male, and was 26-39 years old. (2) Participants read and responded to one instead of two paragraphs as a supposed botcheck (the one about tooth brushing). Assignment to the P-Flipped and Other-P Flipped conditions was the same as in Study 6, namely, participants were told, “We have selected YOU [THE OTHER PARTICIPANT] to choose Heads/Tails and do the virtual coin flip, to determine which of you gets to be in the Reward Condition and which will be in the Punishment Condition.”

In this study we also included an additional 2 conditions. The main purpose behind these conditions was to explore the effects of different ways of assigning participants to the negative outcome. In the random condition, participants were simply randomly assigned to the good or bad outcome, supposedly on the basis of their mTurkID and whether the day of the month happened to be an even or odd number. Participants in this condition were told: “Who will be in the Reward Condition and who will be in the Punishment Condition will be determined in an arbitrary way, namely by whether the day of the month today is even or odd and by a random feature of your and the other participant's mTurk ID number.” The instructions further explained that the computer would check whether the day of the month is odd or even, and then the mTurk IDs of both participants. If the day of the month is even, it was explained, then the participant who has fewer letters in their mTurk ID would get to be in the reward condition and the other participant would be in the punishment condition. If the day of the month is odd, then the participant who has more letters in their mTurk ID would get to be in the reward condition and the other participant would be in the punishment condition. Participants saw a throbber for 8 seconds while the computer was supposedly making this calculation, and saw the following: “The day of the month is an ODD NUMBER [April 5th], and the OTHER PARTICIPANT has MORE letters in their mTurk ID. Therefore, he/she will be in the Reward condition, and you will be in the Punishment Condition”.

In the competition condition, we had participants compete for who would be in the reward condition and who would be in the punishment condition. The competition was described as follows: “In a moment you will see the screen countdown seconds from 5 (e.g., 5 - 4 - 3 - 2 - 1). The next screen will then show a captcha with embedded letters; that is, a graphic that displays four numbers. You should type those numbers in as fast as you can”. Whoever typed in the words the fastest, participants learned, would get to be in the reward condition. When they advanced the screen they saw the countdown, then a captcha with four numbers in it. After they typed in the numbers and advanced the screen, they saw a throbber for 10 seconds while the computer was supposedly determining the winner. They then read: “The other participant [ND86] entered the correct numbers the quickest. Therefore, he will be in the REWARD condition, while you will be in the PUNISHMENT condition”.

We note that in addition to participants who flipped or didn’t flip the coin, the demographics reported above include participants from the random and competition conditions. These results, however, will be reported in a separate article as this present line of research focuses purely on the illusion of unfairness due to the seemingly controllable coin flip procedure

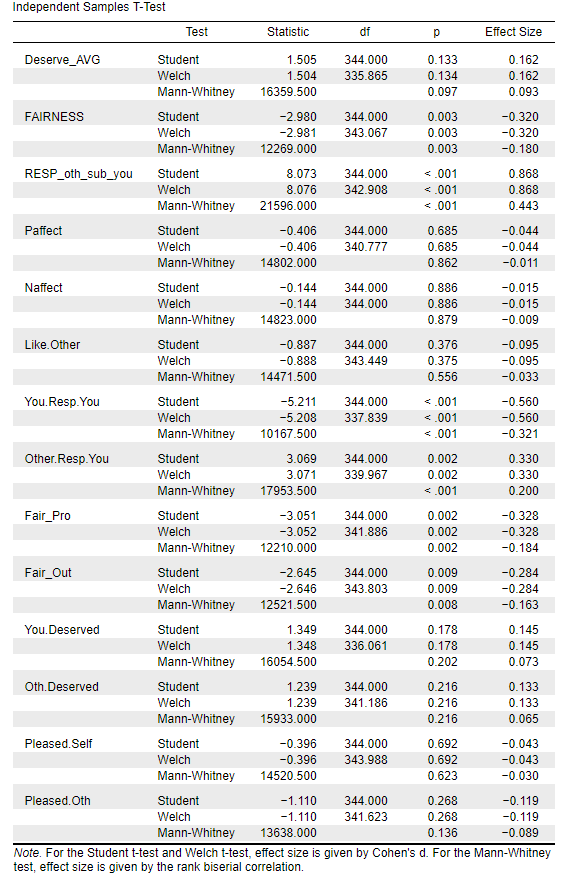
**Results**

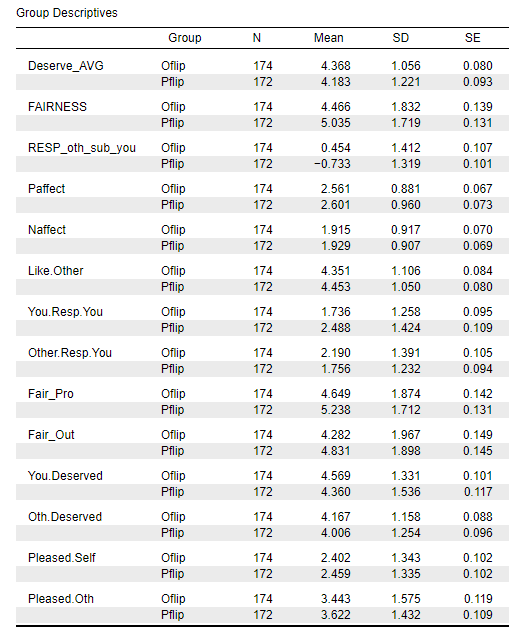
The level of reported suspicion at the end of the study was similar to the previous studies conducted on mTurk: When those in the P-Flipped and Other-P Flipped condition were asked whether they thought the virtual coin spin was random, 29% responded “not at all.” When asked whether they thought there was another participant taking part at the same time, 23% responded “not at all.” When asked how much they believed that they would be in the punishment condition, 5% said not at all. As with the previous mTurk studies, however, when we dropped suspicious participants from the analyses, the results changed little. We again opted to retain participants in the analysis regardless of their suspicion ratings.

The following dependent variables were averaged: Fairness ratings 2 items (.900), deservingness 2 items (.650), negative affect 5 items (.859), and positive affect 5 items (.850). Other dependent measures (e.g., liking for the other participant) were single items.

The descriptive results independent samples *t*-tests comparing the P-Flipped vs Other-P Flipped conditions are reported below in Table S13.

## Table S13: Descriptives and t-test results Study 7





|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Study 7: Descriptives for All Four Conditions** | | | | | | | | |
|  | | N | Mean | Std. Deviation | Std. Error | 95% Confidence Interval for Mean | |
| Lower Bound | Upper Bound |
| Deserve\_AVG | 1.00 OFlips | 174 | 4.3678 | 1.05611 | .08006 | 4.2098 | 4.5258 |
| 2.00 PFlips | 172 | 4.1831 | 1.22111 | .09311 | 3.9993 | 4.3669 |
| 3.00 mTurkID | 165 | 4.3879 | 1.10756 | .08622 | 4.2176 | 4.5581 |
| 4.00 Competition | 166 | 3.7500 | 1.56646 | .12158 | 3.5099 | 3.9901 |
| Total | 677 | 4.1743 | 1.27443 | .04898 | 4.0781 | 4.2705 |
| FAIRNESS | 1.00 OFlips | 174 | 4.4655 | 1.83226 | .13890 | 4.1914 | 4.7397 |
| 2.00 PFlips | 172 | 5.0349 | 1.71899 | .13107 | 4.7762 | 5.2936 |
| 3.00 mTurkID | 165 | 4.8061 | 1.83096 | .14254 | 4.5246 | 5.0875 |
| 4.00 Competition | 166 | 4.6416 | 1.92263 | .14922 | 4.3469 | 4.9362 |
| Total | 677 | 4.7363 | 1.83498 | .07052 | 4.5979 | 4.8748 |
| RESP\_oth\_sub\_you | 1.00 OFlips | 174 | .4540 | 1.41244 | .10708 | .2427 | .6654 |
| 2.00 PFlips | 172 | -.7326 | 1.31944 | .10061 | -.9311 | -.5340 |
| 3.00 mTurkID | 165 | -.0848 | .61887 | .04818 | -.1800 | .0103 |
| 4.00 Competition | 166 | -.8012 | 1.42370 | .11050 | -1.0194 | -.5830 |
| Total | 677 | -.2866 | 1.34363 | .05164 | -.3880 | -.1852 |
| Paffect | 1.00 OFlips | 174 | 2.5609 | .88132 | .06681 | 2.4290 | 2.6928 |
| 2.00 PFlips | 172 | 2.6012 | .96038 | .07323 | 2.4566 | 2.7457 |
| 3.00 mTurkID | 165 | 2.4655 | .84367 | .06568 | 2.3358 | 2.5951 |
| 4.00 Competition | 166 | 2.4518 | 1.04003 | .08072 | 2.2924 | 2.6112 |
| Total | 677 | 2.5211 | .93422 | .03590 | 2.4506 | 2.5916 |
| Naffect | 1.00 OFlips | 174 | 1.9149 | .91696 | .06951 | 1.7777 | 2.0521 |
| 2.00 PFlips | 172 | 1.9291 | .90733 | .06918 | 1.7925 | 2.0656 |
| 3.00 mTurkID | 165 | 1.9745 | .94228 | .07336 | 1.8297 | 2.1194 |
| 4.00 Competition | 166 | 1.9361 | .87968 | .06828 | 1.8013 | 2.0710 |
| Total | 677 | 1.9383 | .91005 | .03498 | 1.8696 | 2.0069 |
| Like.Other | 1.00 OFlips | 174 | 4.35 | 1.106 | .084 | 4.19 | 4.52 |
| 2.00 PFlips | 172 | 4.45 | 1.050 | .080 | 4.30 | 4.61 |
| 3.00 mTurkID | 165 | 4.35 | .922 | .072 | 4.20 | 4.49 |
| 4.00 Competition | 166 | 4.29 | 1.155 | .090 | 4.11 | 4.47 |
| Total | 677 | 4.36 | 1.062 | .041 | 4.28 | 4.44 |
| You.Resp.You | 1.00 OFlips | 174 | 1.74 | 1.258 | .095 | 1.55 | 1.92 |
| 2.00 PFlips | 172 | 2.49 | 1.424 | .109 | 2.27 | 2.70 |
| 3.00 mTurkID | 165 | 1.55 | 1.107 | .086 | 1.38 | 1.72 |
| 4.00 Competition | 166 | 3.63 | 1.271 | .099 | 3.44 | 3.83 |
| Total | 677 | 2.35 | 1.507 | .058 | 2.23 | 2.46 |
| Other.Resp.You | 1.00 OFlips | 174 | 2.19 | 1.391 | .105 | 1.98 | 2.40 |
| 2.00 PFlips | 172 | 1.76 | 1.232 | .094 | 1.57 | 1.94 |
| 3.00 mTurkID | 165 | 1.47 | 1.051 | .082 | 1.31 | 1.63 |
| 4.00 Competition | 166 | 2.83 | 1.360 | .106 | 2.62 | 3.04 |
| Total | 677 | 2.06 | 1.363 | .052 | 1.96 | 2.16 |
| Fair\_Pro | 1.00 OFlips | 174 | 4.65 | 1.874 | .142 | 4.37 | 4.93 |
| 2.00 PFlips | 172 | 5.24 | 1.712 | .131 | 4.98 | 5.50 |
| 3.00 mTurkID | 165 | 5.00 | 1.825 | .142 | 4.72 | 5.28 |
| 4.00 Competition | 166 | 4.69 | 1.959 | .152 | 4.39 | 4.99 |
| Total | 677 | 4.90 | 1.856 | .071 | 4.76 | 5.04 |
| Fair\_Out | 1.00 OFlips | 174 | 4.28 | 1.967 | .149 | 3.99 | 4.58 |
| 2.00 PFlips | 172 | 4.83 | 1.898 | .145 | 4.55 | 5.12 |
| 3.00 mTurkID | 165 | 4.61 | 1.983 | .154 | 4.31 | 4.92 |
| 4.00 Competition | 166 | 4.59 | 1.960 | .152 | 4.29 | 4.89 |
| Total | 677 | 4.58 | 1.958 | .075 | 4.43 | 4.73 |
| You.Deserved | 1.00 OFlips | 174 | 4.57 | 1.331 | .101 | 4.37 | 4.77 |
| 2.00 PFlips | 172 | 4.36 | 1.536 | .117 | 4.13 | 4.59 |
| 3.00 mTurkID | 165 | 4.59 | 1.297 | .101 | 4.39 | 4.79 |
| 4.00 Competition | 166 | 3.91 | 1.792 | .139 | 3.64 | 4.18 |
| Total | 677 | 4.36 | 1.522 | .058 | 4.25 | 4.48 |
| Oth.Deserved | 1.00 OFlips | 174 | 4.17 | 1.158 | .088 | 3.99 | 4.34 |
| 2.00 PFlips | 172 | 4.01 | 1.254 | .096 | 3.82 | 4.19 |
| 3.00 mTurkID | 165 | 4.18 | 1.211 | .094 | 4.00 | 4.37 |
| 4.00 Competition | 166 | 3.59 | 1.534 | .119 | 3.36 | 3.83 |
| Total | 677 | 3.99 | 1.314 | .051 | 3.89 | 4.09 |
| Pleased.Self | 1.00 OFlips | 174 | 2.40 | 1.343 | .102 | 2.20 | 2.60 |
| 2.00 PFlips | 172 | 2.46 | 1.335 | .102 | 2.26 | 2.66 |
| 3.00 mTurkID | 165 | 2.25 | 1.246 | .097 | 2.06 | 2.44 |
| Competition | 166 | 2.33 | 1.453 | .113 | 2.10 | 2.55 |
| Total | 677 | 2.36 | 1.346 | .052 | 2.26 | 2.46 |
| Pleased.Oth | 1.00 OFlips | 174 | 3.44 | 1.575 | .119 | 3.21 | 3.68 |
| 2.00 PFlips | 172 | 3.62 | 1.432 | .109 | 3.41 | 3.84 |
| 3.00 mTurkID | 165 | 3.35 | 1.484 | .116 | 3.12 | 3.57 |
| 4.00 Competition | 166 | 3.10 | 1.600 | .124 | 2.85 | 3.34 |
| Total | 677 | 3.38 | 1.532 | .059 | 3.26 | 3.50 |

# Study 8

**Overview**

Participants learned that who got to call heads/tails and flip the coin would be determined randomly (by how many letters were in the participants’ mTurk ID). Half of the participants were randomly assigned to a consent condition, in which they signed a consent form acknowledging that the coin flip assignment was random and arbitrary. The design was thus a 2 (Who Flipped: Participant or their Opponent) x Consent (no consent, consent) design.

**Method**

**Participants.** Participants were recruited from Cloud Research. We set an approval rating of at least 85% and the number of hits completed to 50 - 10,000. Of the 1,101 participants who began the study, 390 dropped out or were dropped before being assigned to condition, for the same reasons as the other mTurk studies. Of the 711 who completed the dependent measures, 342 identified as female, 368 identified as male, and 1 identified as other. Participants indicated that they were in the following age ranges: 18-25 (58), 26-39 (342), 40-59 (235), and 60 or older (76). Participants self-identified their race as follows: 496 White, 118 Black, 43 Asian, 26 Latino/a, 1 American Indian/Alaskan Native, and 16 Multiracial. As discussed below, we dropped 82 participants from the analyses who failed one or more manipulation checks. Of the remaining 629 participants, 315 identified as female, 313 identified as male, and 1 identified as other. These participants indicated that they were in the following age ranges: 18-25 (51), 26-39 (280), 40-59 (225), and 60 or older (73), and self-identified their race as follows: 471 White, 67 Black, 40 Asian, 26 Latino/a, 1 American Indian/Alaskan Native, and 16 Multiracial.

**Procedure.** The procedure was identical to the previous mTurk studies except for the manner in which it was determined who would get to call heads/tails and flip the coin. All participants were told that the assignment would be determined by who had more letters in their mTurk ID. (Note that Study 9 compared this procedure to the standard coin flip assignment.) The instructions presented to participants for these two conditions were the following:

“An arbitrary way to decide who gets to choose Heads/Tails and flip the coin is to see who has more letters in their mTurk ID (as opposed to numbers). Whoever has more letters will be the one to call Heads/Tails and flip the coin. Our computer will compare your mTurk IDs and give the answer on the next page.”

Participants were then assigned to flipping conditions with the instructions below:

* P-Flipped condition: “YOU have more letters in your mTurk ID. Therefore, you will get to call Heads or Tails and flip the virtual coin, to determine who will be in the reward condition and who will be in the punishment condition.”
* Other-P Flipped condition: “THE OTHER PARTICIPANT has more letters in their mTurk ID. Therefore, he/she will get to call Heads or Tails and flip the virtual coin, to determine who will be in the reward condition and who will be in the punishment condition.”

We also manipulated whether participants gave their consent for this assignment procedure. In the consent condition, after learning that who flipped would be determined by the participants who had more letters in their mTurk ID, participants read, “Before we proceed, we want to make sure you agree that this way of deciding who gets to call Heads/Tails and flip the coin is random.” They were asked to read and agree to the following three statements:

● I understand that the number of letters in a person's mTurk ID is arbitrary and random.

● I understand it doesn't really matter who gets to call Heads/Tails and flip the coin, because the coin flip is random.

● I'm OK with me or the other participant calling Heads/Tails and flipping the coin, however it turns out.

Participants were then asked to “sign with your initials to indicate that you understand and consent to this way of determining who gets to call Heads/Tails and flip the coin.” All participants gave their approval, even though they were free to move forward without signing. In the no consent condition participants were not asked to consent to the assignment protocol. From this point on the procedure was identical to Study 7.

**Results**

The level of reported suspicion at the end of the study was similar to the previous studies conducted on mTurk: When asked whether they thought the virtual coin spin was random, 22% responded “not at all.” When asked whether they thought there was another participant taking part at the same time, 23% responded “not at all.” When asked how much they believed that they would be in the punishment condition, 8% said not at all. As with the previous mTurk studies, however, when we dropped suspicious participants from the analyses, the results changed little. We again opted to retain participants in the analysis regardless of their suspicion ratings.

As preregistered, we dropped 82 participants who either failed to recall corrected who got to flip the coin or they would be in the punishment or reward condition. Somewhat more of the people who failed the manipulation checks were in the Other-P Flipped condition (*n* = 51) than the P-Flipped condition (*n* = 31), *X2* (*df*  =1, *N* = 711) = 3.20, *p* = .018.

We computed the same scales as reported in Studies 5 and 6, which had these alphas: Fairness ratings (.903), deservingness (.721), how pleased participants were (.746), negative affect (.903), and positive affect (.875). Other dependent measures (e.g., liking for the other participant) were single items.

As seen in Table S14, a series of 2 (Who Flipped) x 2 (Consent) ANOVAs revealed a significant main effect of Who Flipped on most of the measures, demonstrating the standard illusion of unfairness. Was the illusion reduced in the consent condition? As it happened, only on ratings of fairness (see Table S14). The Who Flipped x Consent interaction was significant only on this measure, reflecting the fact that in the no consent condition, participants thought the procedure was significantly less fair when the other participants flipped the coin, whereas in the consent condition, there was no significant difference in fairness ratings based on who flipped the coin.

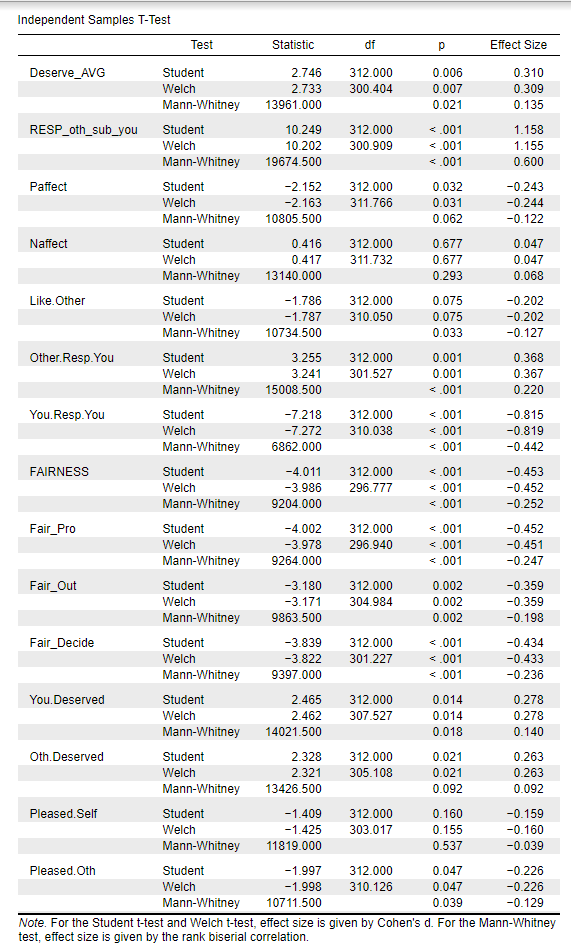
Thus, with the exception of the question about fairness, there was scant evidence that emphasizing the randomness of the selection process, and giving participants a voice by asking them to consent to the procedure, reduced the illusion of unfairness. Why did participants in the consent condition moderate their responses on the fairness question? One possibility is experimental demand. Participants in the consent condition may have begrudgingly admitted that the procedure was fair, after agreeing with such statements as, “I understand it doesn't really matter who gets to call Heads/Tails and flip the coin, because the coin flip is random.” Nonetheless they still felt less responsible for the outcome when the other person flipped the coin, and were less pleased with the outcome (see Table S14).

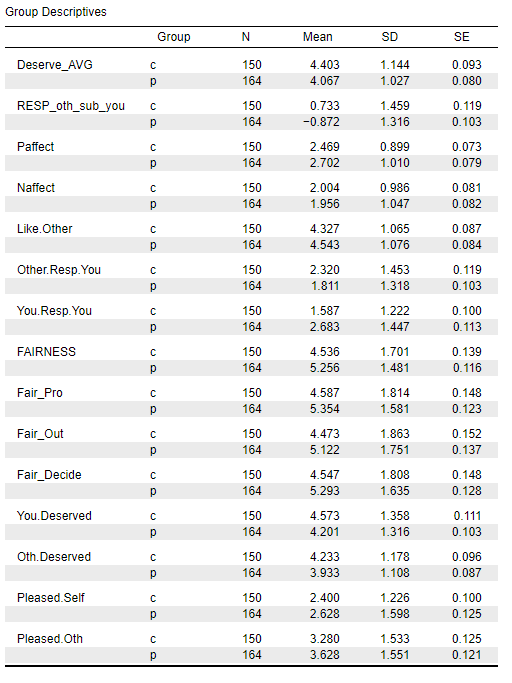
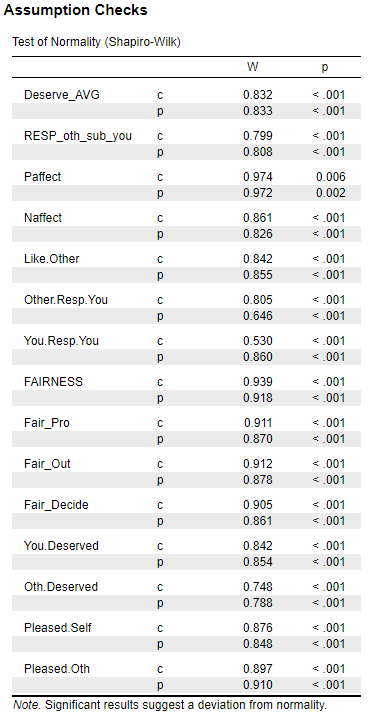
|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Table S14: Study 8: Effects of Who Flipped and Consenting to How it Would Be Determined Who Flipped | | | | | | | | | |
|  | | | | | | | | | |
| Variables | Means (*SD*s) by Condition | | | | ANOVA: *F*(1, 625) | | | Simple Effect of Who Flipped: *F*(1, 625) | | | |
|  | No Consent | | Consent | |
|  | P-Flipped | Other P  Flipped | P-Flipped | Other P  Flipped | Who  Flipped | Reflection Form | Interaction | | No Consent | | Consent |
| Fairness | 5.26 (1.48) | 4.54 (1.70) | 5.25 (1.62) | 5.19 (1.70) | 8.97\*\* | 6.30\* | 6.54\* | | 0.10 | | 15.45\*\*\* |
| You Responsible | 2.68 (1.45) | 1.59 (1.22) | 2.65 (1.57) | 1.53 (1.15) | 103.16\*\*\* | 0.16 | 0.01 | | 52.43\*\*\* | | 50.74\*\*\* |
| Other P Responsible | 1.81 (1.32) | 2.32 (1.45) | 1.86 (1.40) | 2.16 (1.37) | 13.61\*\*\* | 0.25 | 0.82 | | 3.86ǂ | | 10.59\*\* |
| Undeservingness of Self Getting Punishment | 4.20 (1.32) | 4.57 (1.36) | 4.42 (1.42) | 4.54 (1.20) | 5.40\* | 0.75 | 1.40 | | 0.65 | | 10.85\* |
| Undeservingness of Other P Getting Reward | 3.93 (1.11) | 4.23 (1.18) | 4.09 (1.28) | 4.10 (1.08) | 2.80 | 0.01 | 2.41 | | 0.01 | | 5.21\* |
| Undeserving Average | 4.07 (1.03) | 4.40 (1.14) | 4.25 (1.21) | 4.32 (1.02) | 5.20\* | 0.33 | 2.35 | | 0.28 | | 7.30\*\* |
| Liking | 4.54 (1.08) | 4.33 (1.07) | 4.48 (1.02) | 4.43 (1.02) | 2.35 | 0.06 | 1.10 | | 0.12 | | 3.34ǂ |
| Pleased with Own Condition Assignment | 2.63 (1.60) | 2.40 (1.23) | 2.83 (1.63) | 2.45 (1.17) | 7.09\*\* | 1.17 | 0.43 | | 5.53\* | | 2.00 |
| Pleased with Other Participant’s Condition Assignment | 3.63 (1.55) | 3.28 (1.53) | 3.47 (1.58) | 3.40 (1.44) | 3.02ǂ | 0.03 | 1.25 | | 0.19 | | 4.06\* |
| Positive Affect | 2.70 (1.01) | 2.47 (0.90) | 2.59 (1.06) | 2.53 (0.97) | 3.51ǂ | 0.11 | 1.14 | | 0.32 | | 4.34\* |
| Negative Affect | 1.96 (1.05) | 2.00 (0.99) | 2.00 (1.01) | 1.95 (1.05) | 0.00 | 0.00 | 0.42 | | 0.24 | | 0.18 |

ǂ = *p* < .10, \* *p* < .05. \*\**p* < .01, \*\*\**p* < .005

Note. Planned contrast = Other-P flip, No Reflection = -3, P-Flipped, No Reflection. = 1, Other-P flip, Reflection = 1, P-flip, Reflection =1.

## Table S15: T-Test Results of Study 8





# Study 9

**Method**

**Participants.** Participants were recruited from Cloud Research. We limited participation to those with an approval rating of at least 90% and hits completed between 100 - 50,000. Of the 1430 participants who began the study by signing the consent form, 1236 were assigned to coin flip conditions and 1216 completed the dependent measures, only 32 were removed because of failing our manipulation checks asking them which version of the learning ask they received or who got to flip the coin. We were left with a total of 1184. Total gender data for 1177 participants was: 547 identified as male and 630 identified as female. 113 participants were between 18-25, 491 were between 26-39, 428 were between 40-59, 152 were 60 and above. Participants self-identified their race as follows: 937 White, 88 Black, 124 Asian, 61 Latino/a, 5 American Indian/Alaskan Native, 39 more than one race.

We randomly assigned participants to the usual two flip conditions (P-Flipped vs. Other-P Flipped), but instead of only asking the dependent variables (i.e. fairness) after the outcome was determined, we presented participants with our dependent variables at two time points in a repeated measures design: Once, before the outcome was determined (i.e. after the coin was supposedly flipped but before participants knew the outcome) and a second time after revealing the outcome to the participants. The coin flip assignment was the same as in the Experimenter Decided condition of Study 6. The only slight differences were that the coin flip manipulation check stayed the same across conditions, but we removed two questions: “Which side did the coin land on” and “Which condition did you receive (punishment/reward). Whereas the regular condition had the coin flip land on a specific outcome and was followed by the DV instructions “Before you start the category learning task… we would like a baseline... ”. In the before outcome condition, the coin flip spinning was interrupted with the DV instructions “Before you get the outcome of the coin flip and start the category learning task…” the rest of the instructions remained the same.

**Dependent Measures.** The dependent variables in this study were measured on slider scales ranging from 0 – Extremely […] to 100 – Extremely […] because they were collected at two time points and we wanted to minimize the possibility that people recalled their exact answer to the first set of questions. A few measures were modified slightly from the previous studies in order to fit with both the before and after outcome conditions.

Participant liking was measured with the following item: “If you were to meet the other participant, how much do you think you would like or dislike him/her?” Responsibility of self and other were measured with the following items: “How responsible do you feel YOU are for the version of the learning task YOU receive? (i.e., the reward or punishment condition)?” and “How responsible do you feel THE OTHER PARTICIPANT is for the version of the learning task YOU receive? (i.e., the reward or punishment condition)?”

Fairness at both time points was measured using the following two items: “Think about how it was decided whether YOU would be in the reward or punishment condition. How fair or unfair do you think this procedure was?” and “Think about how it was decided whether you or the other participant would get to call heads/tails and flip the coin. How fair or unfair do you think this procedure was?” Outcome fairness could only be measured at time 2 with the following item: “How fair does the outcome of the coin flip feel to you?”

We measured participant and other participant undeservingness at both time points using the following four items: “To what extent do you feel you deserve the reward condition?” and “To what extent do you feel you deserve the punishment condition?”, “To what extent do you feel the other participant deserves the reward condition)?”, and “To what extent do you feel the other participant deserves the punishment condition?

We measured how pleased participants felt about their outcome only at Time 2-- once they found out their outcome--using the following items: “How pleased/displeased are you about the condition YOU are in (i.e., the reward or the punishment condition)?” and “How pleased/displeased are you about the condition the other participant is in (i.e., the reward or the punishment condition)?”

Fairness items: how fair was it who got to flip, and how fair was the process were averaged and used as the fairness DV, the outcome fairness and how pleased questions were moved at the very end of time 2 DVs since they cannot be asked before the outcome is revealed. Note these 3 additional exploratory variables came after the manipulation checks of who flipped and which version of the learning task participants received. Since coded suspicion never made much of a difference on our results, we removed four of the open-ended questions, except general impressions.

**Results**

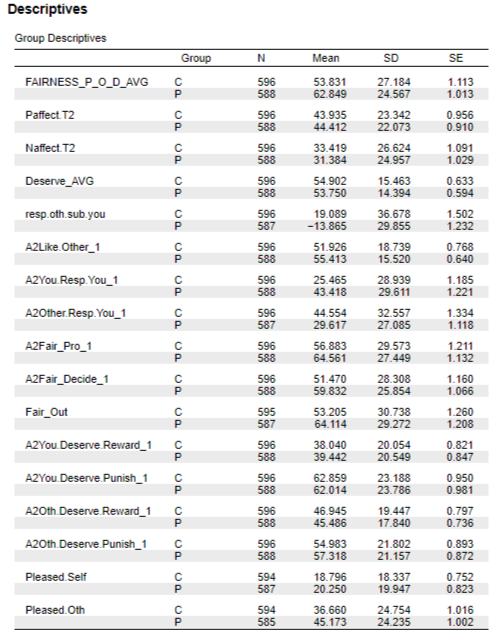
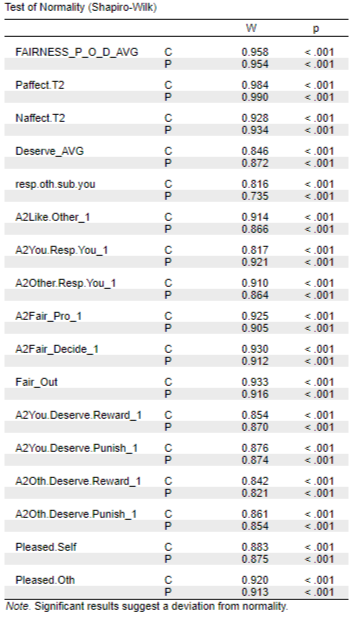
The level of reported suspicion at the end of the study was similar to the previous mTurk studies. When asked whether they thought the virtual coin spin was random, 23.1% responded “not at all.” (107, P-Flipped), (166, Other-P Flipped). When asked whether they thought there was another participant taking part at the same time, 25% responded “not at all.” (139, P-Flipped), (157, Other-P Flipped). When asked how much they believed that they would be in the punishment condition, 3.9% responded “not at all.” (23, P-Flipped), (23, Other-P Flipped).

As with the previous mTurk studies, however, when we dropped suspicious participants from the analyses, the results changed little. We again opted to retain participants in the analysis regardless of their suspicion ratings.

We computed the same scales as reported in past studies, which had these alphas for DVs at time 1 and time 2, respectively: Fairness ratings (.787, .894), negative affect (.838, .885), and positive affect (.784, .876), undeservingness of the actual outcome (.-.149, -.014) and undeservingness of the alternative outcome(-1.85, -.095). Because the reliability of the undeservingness items was so low, we reported each item individually in the results (see main manuscript). We suspect that asking participants about deservingness before and after receiving the results led to drastically different interpretations of the question, which explains why the reliability index was low for these items and therefore led us to analyze them each individually. Other dependent measures (e.g., liking for the other participant) were single items.

The descriptives and independent samples t-tests comparing the P-Flipped vs Other-P Flipped conditions are reported below in Table S16. Note that all dependent variables noted in Table S16 are from Time 2 reports in Study 11. Those were the variables used in the meta-analysis as they were the most similar to the other studies included in the meta-analysis.

## Table S16. Descriptives and *t*-test Results at Time 2 for Study 9



# Study 10

**Method**

**Participants.** Participants were recruited from Cloud Research on Saturday, March 7, 2020. We set an approval rating of at least 85% and the number of hits completed to 50 - 10,000. Of the 947 participants who began the study, 228 dropped out or were dropped before being assigned to condition for the same reasons as the other mTurk studies. Of the 714 who completed the dependent measures, 323 identified as male, 388 identified as female, and 3 identified as other. Participants indicated that they were in the following age ranges: 18-25 (57), 26-39 (312), 40-59 (255), and 60 or older (90). Participants self-identified their race as follows: 560 White, 60 Black, 38 Asian, 22 Latino/a, 3 American Indian/Alaskan Native, and 17 Multiracial. As discussed below, we dropped 90 participants from the analyses who failed one or more manipulation checks. Of the remaining 624 participants, 275 identified as male, 346 identified as female, and 3 identified as other. Participants indicated that they were in the following age ranges:18-25 (51), 26-39 (258), 40-59 (227), and 60 or older (88). These participants self-identified their race as follows: 495 White, 50 Black, 35 Asian, 19 Latino/a, 2 American Indian/Alaskan Native, and 14 Multiracial.

**Procedure.** The procedure was identical to Study 6 except we manipulated how it was determined who would get to flip the coin. In the No Explanation condition (randomly assigned), participants read, “We have selected YOU [THE OTHER PARTICIPANT] to choose Heads/Tails and do the virtual coin flip, to determine which of you gets to be in the Reward Condition and which will be in the Punishment Condition.” In the explanation condition participants learned that who got to call Heads/Tails and flip the coin would be termed by who had more letters in their mTurk ID, as was done in Study 6. The remaining procedure and dependent measures were identical to those in Study 6.

**Results**

Once again there was a fair amount of reported suspicion about various aspects of the procedure. When asked whether they thought the virtual coin spin was random, 29% responded “not at all.” When asked whether they thought there was another participant taking part at the same time, 25% responded “not at all.” When asked how much they believed that they would be in the punishment condition, 4% said not at all. As with the previous mTurk studies, however, when we dropped suspicious participants from the analyses, the results changed little. We again opted to retain participants in the analysis regardless of their suspicion ratings.

As preregistered, we dropped 90 participants who failed either one or both of three manipulation checks: Recalling who got to flip the coin, recalling whether they would be in the punishment or reward condition, and recalling how it was determined who would get to flip the coin. There was little difference in whether those dropped were in the Other-P Flipped condition (*n* = 47) or P-Flipped condition (*n* = 43), *X2* (1, *N* = 714) = .20, *p* = .652.

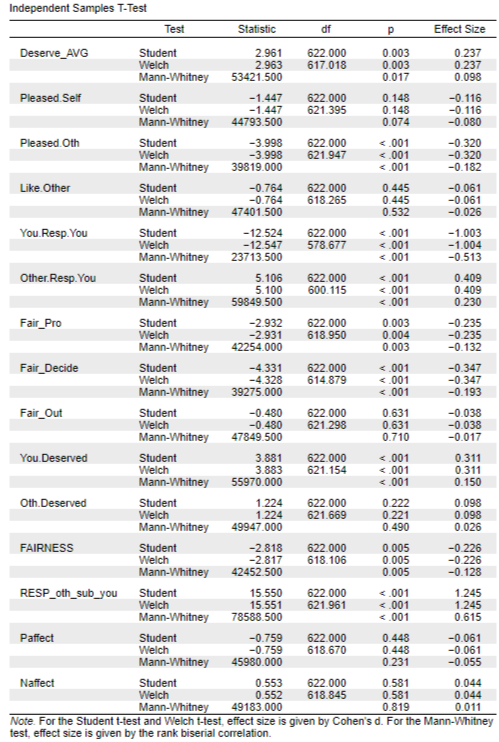
We computed the same scales as reported in previous studies, which had these alphas respectively: Fairness ratings (.90), deservingness (.721), how pleased participants were (.746), negative affect (.903), and positive affect (.875). Other dependent measures (e.g., liking for the other participant) were single items.

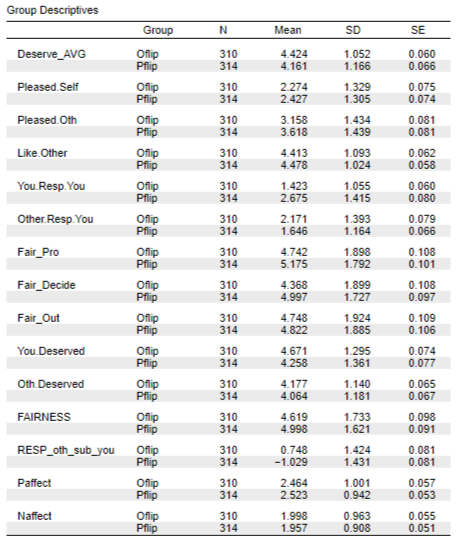
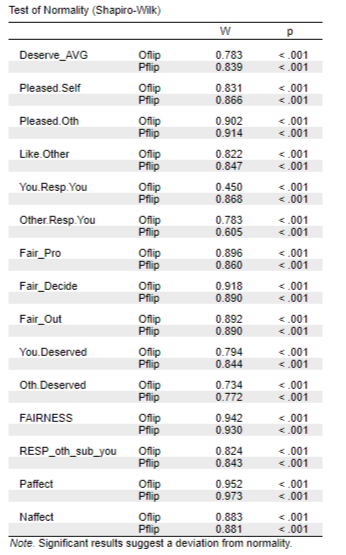
The descriptive results and the independent samples *t*-tests comparing the P-Flipped vs. Other-P Flipped conditions are reported below in Table S17.

The effects of who flipped the coin were very similar to the previous studies. When the other participant flipped, participants reported that the procedure and outcome were less fair, that they were more undeserving of the outcome, and that the other participant was more responsible for the outcome, than when they flipped the coin. These differences were significant on each of these measures in both studies, with the exception of the fairness measure, where *p* = .079. Once again the effect sizes ranged from small to large (*d*s = .141 to 1.247). Consistent with the previous findings, participants in the Other-P Flipped condition reported that they were significantly less pleased with the outcome than did participants in the P-Flipped condition, but there were no significant differences on negative or positive affect.

As reported in the paper, there was little effect of whether an explanation was provided for who got to flip the coin. There were no significant effects of how the flip was determined, nor were any of the Who Flipped x How Determined interactions significant. The only exception was a main effect of How Determined on negative affect, *F*(1, 620) = 4.394, *p* = .036, partial eta sq = .007, reflecting the fact that those in the mTurk ID condition reported slightly more negative affect (*M* = 2.05, *SD* = 1.00) than did those the Experimenter Decided condition (*M* = 1.90, *SD* = .86). The Who Flipped x How Determined interaction was not significant, *F*(1, 620) = .151, *p* = .698, partial eta sq < .001. On all of the other dependent measures shown in Table S17, the How Determined main effect was nonsignificant, *F*s(1, 620) < 1.23, as was the Who Flipped x How Determined interaction, *F*s(1, 620) < .68. Therefore, to be consistent with the previous studies, Table S17 displays the means in the P-Flipped and Other-P Flipped conditions collapsed across the How Determined manipulation.

## Table S17. Descriptives and *t*-test Results Study 10





**Bayesian Interval-Null testing Results.** We conducted exploratory Bayesian Interval-Null testing (equivalence testing) to determine how much evidence our data provide in favor of the null, that is, that the difference between the explanation conditions was negligible. Specifically, we tested whether the explanation manipulation had no effect on participants who flipped the coin (P-Flipped) as well as on participants who did not flip the coin (Other-P Flipped). We set our equivalence region to -0.5 – 0.5, in order to be conservative in our estimates, and used the default Cauchy prior with a scale of 0.707. We tested the degree to which the data support the hypothesis that the parameter lies inside versus outside the equivalence region, and found that, for participants who did not flip the coin, the non-overlapping-hypothesis Bayes factor in favor of the interval-null ranged from 3680 (negative affect) to 217000 (other responsibility). We also tested the same hypothesis when participants flipped the condition (P-Flipped conditions), the non-overlapping-hypothesis Bayes factor in favor of the interval-null ranged from 471 (negative affect) to 241000 (fairness outcome). Overall, these results demonstrate extremely strong evidence for the null, hypothesis when comparing participants in the explanation condition with those who did not receive the explanation.

## Table S18. Results Comparing Explanation Condition with No Explanation Condition Using only Other-P Flipped Condition

## Table S19. Results Comparing Explanation Condition with No Explanation Condition using Only Participant Flip Condition

# Study 11

**Method**

**Overview.** This study, run on Cloud Research, tested a dual process approach to the illusion of unfairness. We hypothesized that the illusion would be exacerbated by having participants make quick intuitive judgments while assessing the fairness of the coin flip procedure. Alternatively, asking participants to reflect before making their fairness judgment might reduce the illusion.

**Procedure.** We used the standard coin flip paradigm that in which participants were simply told was selected to flip the coin.[[1]](#footnote-1) All participants were assigned to the Other-P Flip condition and were randomly assigned to one of three response conditions: A Deliberative condition, an Intuitive condition and a Control Condition (in all conditions, the other participant flipped the coin).

In the Intuitive Condition participants were told:

IMPORTANT: We are interested in your immediate “gut reactions” to the following questions. Please give the response that first comes to mind, without thinking about it too much. To encourage you to give your gut response, you will only have 8 seconds to answer each question. If you haven’t responded after 8 seconds, the program will automatically advance to the next question. So, please be ready to read the question and respond quickly with your gut response. (Pretesting indicated that this procedure would induce participants to respond more quickly than they normally would.)

In the Deliberative Condition, participants were directed to answer the dependent variables with the following instructions:

IMPORTANT: We are interested in your logical, reasoned responses to the following questions. Please reflect on each question carefully and give what you think is the most reasoned response. To encourage you to give your logical, reasoned response, there will be an 8 second delay between the time you read the question and the opportunity to respond to it. Please spend that time reflecting on the question and thinking about your most reasoned response.

In the Control Condition participants did not receive any instructions about how fast or slow to respond. All participants then received these instructions, so that they were familiar with the response scales in advance:

IMPORTANT: There will be a total of 10 questions. Each one will be answered on a 5-point scale that with these responses:

Not at All Slightly Somewhat Moderately Very Much

***Dependent measures.*** Due to the timing manipulations in the deliberative and intuitive conditions, we set all the dependent variables on a 5-point scale to increase ease of responding for participants. We also used a paired down version of the dependent variables to decrease participants’ response burden (see below).

One item assessed how displeased/pleased they were with the condition they had been assigned (i.e., “How pleased are you about the condition you are in (i.e., the reward or the punishment condition)?”), on a 5-point scale (*1 = Not at all, 5 = Very Much*).

Participants then rated the extent to which they were happy and angry (i.e., “How Happy do you feel right now?”) on 5-point scales (*1 = Not at all, 5 = Very much*). Participants then rated how much they liked the other participant (i.e., “How much do you like the other participant?” (*1 = Not at all, 5 = Very Much*).

Next, participants answered three questions about how fair they found: the coin flip assignment, the procedure, the process, and the outcome (i.e., “How fair was it that the other person got to call heads/tails and flip the coin?; Think about how it was decided whether you would be in the reward or punishment condition. How fair do you think this procedure was? How fair does the outcome of the coin flip feel to you?”, on 5-point scales (*1 = Not at all, 5 = Very Much*).

Subsequently, participants rated how responsible they thought they and the participant were for the fact that they were assigned to the punishment condition (i.e., “How responsible are you for the outcome you received (reward or punishment condition)?” and “How responsible do you feel the other participant is for the version of the learning task you received? (i.e., the reward or punishment condition)?”

Finally, participants rated how deserving they thought it was that they and the other participant received the outcomes they did (i.e., “How much do you deserve the outcome you received (reward or punishment condition)?)” on a 5-point scale (*1 = not at all, 5 = Very Much*).

**Results**

The level of reported suspicion at the end of the study was similar to the previous mTurk studies. When asked whether they thought the virtual coin spin was random, 29.6% responded “not at all”. When asked how much they believed that they would be in the punishment condition, 4.7% responded “not at all.”. When asked whether they thought there was another participant taking part at the same time, 26.7% responded “not at all.” Once again the results changed little when suspicious participants were dropped from the analyses, so, as pre-registered, we included all participants.

We computed the same three item fairness scales as reported in previous studies, which had an alpha reliability rating of .886., negative affect (anger), positive affect (happy) and Deservingness (self), as well as other dependent measures (e.g., liking for the other participant) were single items.

This study was not included in the meta-analysis because there were no Participant Flip conditions. The full results of this study are presented in the main manuscript.

# Study 12

**Method**

**Participants**. For this field study we recruited participants from a library on campus as and in front of a dining commons. One hundred twenty-eight participants completed the study. Forty-five identified as male, 82 identified as female, 1 identified as gender fluid. Participants’ age ranged from 18-65 years old with a mean of 21.5 years. Participants self-identified their race as follows: 76 White, 9 African American, 37 Asian, 9 Hispanic, 0 Native American, 1 Pacific Islander, 1, Mixed, 1 other. All but one participant correctly reported having won/lost as well as having flipped the coin or not. We included all participants in our analyses because when we dropped: 1 participant who mistakenly reported having won when they in fact lost, 9 participants who reported having completed one of the lab studies regarding the illusion of unfairness, 4 participants who re-flipped the coin because of an issue, 4 participants who flipped or called their preference too early, 3 participants who expressed not wanting to flip, we found no major differences compared to including them in the final dataset. We therefore decided to keep all participants.

**Procedure.** The procedure was similar to that of the first field study except that the participant was competing for the prize not with another participant, but with the experimenter to determine whether they would be compensated with a prize for filling out a survey. That is, instead of competing against another participant who was equally deserving of the reward, participants competed against the experimenter. If participants lost the flip they thus ended up helping the experimenter (by filling out a survey) for no compensation. If they won the flip they received appropriate compensation for their participation. Winning or losing was thus psychologically different than all the other studies. (At the end of the study, after being debriefed, all participants received the prize.) Specifically, the experimenter said, “OK, let’s flip the coin” (while showing both sides of the coin to the participant). “If you win, you get the prize, if I win, I get to keep the prize. So, it’s kind of like a contest between us. Sometimes I let the participant flip the coin and other times I go ahead and flip the coin.”

In the Participant Flip condition (randomly assigned) the experimenter then said, *“You* can flip the coin. First, please choose Heads or Tails . . .You’re going to flip the coin like this with your thumb and index finger and let the coin fall on the floor. Please do not flip it just yet.” In the Experimenter Flip condition the experimenter said, “I can flip the coin. First, I’ll choose Heads or Tails. Let’s see ... I want tails. Now I’ll flip the coin like this [shows them]. I’ll flip it with my thumb and index finger and let the coin fall on the floor.” If the participant won the coin toss the experimenter said, “OK, you win the prize. Go ahead and choose one of the drinks or snacks.” If the experimenter won, they said, “OK, I get to keep the prize.” Participants then completed the dependent variables on a tablet.

The following dependent variables were averaged: 3 Fairness items (process, outcome and how decided) ratings (.781), 5 items negative affect (.804), and 5 items positive affect (.648), two items for deservingness (.663), 2 items for liking (.753). The other dependent measures (e.g., responsibility and guilt) were single items.

**Results: Lose Condition**

The descriptive results and independent samples t-tests comparing the P-Flipped vs Other-P Flipped conditions are reported below in Table S20.

## Table S20. Descriptives and *t*-test results Study 12 (Lose Condition)

A table of numbers and text

Description automatically generated

A table of text with numbers

Description automatically generatedA table of numbers and text

Description automatically generated

**Note. Other-P Flipped condition = 0, P-Flipped condition = 1**

**Results: Win Condition**

The descriptive results and independent samples t-tests comparing the P-Flipped vs Other-P Flipped conditions are reported below in Table S21.

## A table with numbers and letters Description automatically generatedTable S21. Descriptives and *t*-test results Study 12 (Win Condition)

## A screenshot of a computer Description automatically generated

# Moral Luck

Our results also speak to the philosophical literature on “moral luck,” which refers to the case in which people are held morally accountable for acts that are not entirely under their control. An example is the case of two drunk drivers, one of whom strikes a pedestrian (who ran into the middle of the road) while the other makes it home without incident. Most people would judge the first driver more harshly, even though the presence or absence of pedestrians was not controllable by either driver (Nagel, 1979; Williams, 1981). Philosophers have noted a paradox with the concept of moral luck; on the one hand, it seems unfair to blame people for things they could not control, but people rarely, if ever, have complete control over a situation, even when they commit acts that most of us would condemn (Hartman, 2017; Nelkin, 2023). In our studies, participants exhibited a strong version of lay moral luck: They believed that the person who flipped the coin was more responsible for a negative outcome, even though the outcome was random and uncontrollable.

Hartman, R. J. (2017) *In defense of moral luck: Why luck often affects praiseworthiness and blameworthiness*. New York: Routledge. <https://doi.org/10.4324/9781315232041>

Nagel, T. (1979). *Mortal questions*, New York: Cambridge University Press. <https://doi.org/10.1017/s003441250001475x>

Nelkin, Dana K., "Moral Luck", The Stanford Encyclopedia of Philosophy (Spring 2023 Edition), Edward N. Zalta & Uri Nodelman (eds.), URL = <https://plato.stanford.edu/archives/spr2023/entries/moral-luck/>.Orlogg, C. (2019, Sept. 19). Portland penny. *Oregon Encyclopedia.* Retrieved Nov. 2, 2021 from: <https://www.oregonencyclopedia.org/articles/portland_penny/#.YYGeE2DMKUk>

Williams, B. (1981). *Moral luck*, Cambridge: Cambridge University Press. https://doi.org/[10.1017/cbo9781139165860](https://doi.org/10.1017/cbo9781139165860)

# Correlation Matrices between Outcome Variables in All Studies

The spreadsheet is found in the following online repository: <https://researchbox.org/3507&PEER_REVIEW_passcode=PTLHHH>

1. We ran one study prior to this one, but there was a problem with the instructions about the coin flip assignment, whereby part of the instructions was not displayed due to survey logic error. Participants weren’t told: “We have selected **THE OTHER PARTICIPANT** [ND8] to choose Heads/Tails and do the virtual coin flip, to determine which of you gets to be in the Reward Condition and which will be in the Punishment Condition.” Instead, they were immediately told; “If the other participant [ND8] WINS the toss, he will be in the Reward Condition and you will be in the Punishment Condition. If the other participant [ND8] LOSES the toss, you will be in the Reward Condition and he will be in the Punishment Condition.”, followed by; “Please Wait While the Other Participant Selects Heads or Tails ”. Because it was not made explicit who was selected to flip the coin we disregarded the first study, we launched the second study (pre-registered) and report the results of the second study which had the correct instructions. [↑](#footnote-ref-1)