Corrigendum

Psychological Science
2023, Vol. 34(1) 137–138

DOI: 10.1177/09567976221139496
www.psychologicalscience.org/PS

Corrigendum: Causal Inference About Good and Bad Outcomes

Original article: Dorfman, H. M., Bhui, R., Hughes, B. L., & Gershman, S. J. (2019). Causal inference about good and bad outcomes. *Psychological Science, 30*(4), 516–525. https://doi.org/10.1177/0956797619828724

Because of a data-entry error, some t statistics, confidence intervals (CIs), and effect sizes were incorrect. The authors have completed a thorough re-analysis of the data, and the overall conclusions expressed in the article were not affected by the errors; however, several results reported in Experiments 1 and 2 did change. The corrected text is as follows:

Experiment 1 Results

Behavioral analyses

As a preliminary manipulation check, we verified that participants' beliefs about hidden causes varied with the outcome valence in a condition-specific manner (Fig. 2a). Participants were more likely to believe that a hidden cause resulted in negative outcomes, as opposed to positive outcomes, overall, $t(71) = 4.71$, $p < .001$, $d = 0.55$, 95% CI = [0.08, 0.20]. Importantly, participants were more likely to believe that the hidden agent had intervened after negative than after positive outcomes in the adversarial condition, $t(71) = 17.56$, $p < .001$, $d = 2.07$, 95% CI = [0.54, 0.67], and after positive than after negative outcomes in the benevolent condition, $t(71) = -10.38$, $p < .001$, $d = 1.22$, 95% CI = [−0.49, −0.33]. Participants were also slightly more likely to believe that the hidden agent had intervened after negative outcomes in the neutral condition, $t(71) = 2.22$, $p = .03$, $d = 0.26$, 95% CI = [0.01, 0.22]. We will revisit this effect in the context of our computational model.

Computational modeling

To characterize the effects of outcome valence and agent type on learning, we first fitted a reinforcement-learning model with six separate learning rates. As shown in Figure 3a, participants generally learned more from positive than from negative outcomes across all conditions, $t(71) = 3.10$, $p < .01$, $d = 0.37$, 95% CI = [0.00, 0.01]. By treating the positivity bias in the neutral condition, $t(71) = 2.17$, $p = .03$, $d = 0.25$, 95% CI = [0.00, 0.02], as a participant-specific baseline and subtracting it from the other conditions, we obtained a relative measure of learning rates for the adversarial and benevolent conditions (Fig. 3b), revealing an underlying sensitivity to condition and valence. A 2 (condition: adversarial vs. benevolent) × 2 (valence: positive vs. negative) repeated measures analysis of variance on relative learning rates revealed no significant main effects ($p = .57$ for condition, $p = .84$ for valence) but a significant interaction, $F(1, 71) = 4.91$, $p < .05$. Consistent with our hypothesis, results showed that the learning-rate advantage for positive versus negative outcomes reverses depending on the causal structure of the task.

As expected, the Bayesian model showed a strong interaction between condition and outcome valence (Fig. 3c), a direct consequence of causal inference. To bolster our claim that causal inference predicts the valence-dependent learning-rate asymmetry, we examined the relationship between intervention judgments and learning rates (derived from the Bayesian model). We found that learning rates were significantly lower for trials in which participants believed that the hidden agent intervened, compared with trials in which participants believed that the hidden agent did not intervene, $t(71) = 10.38$, $p < .001$, $d = 1.22$, 95% CI = [0.02, 0.04] (Fig. 4a).

Experiment 2 Results

Behavioral analyses

Results of Experiment 2 replicated those of Experiment 1: Participants believed that the hidden agent caused negative outcomes more often than positive outcomes across all conditions, $t(254) = 6.26$, $p < .001$, $d = 0.39$, 95% CI = [0.08, 0.15]. Participants were significantly more likely to believe that the hidden agent had intervened after negative compared with positive outcomes in the adversarial condition, $t(254) = 43.72$, $p < .001$, $d = 2.74$, 95% CI = [0.68, 0.74], and after positive compared with negative outcomes in the benevolent condition, $t(254) = -21.78$, $p < .001$, $d = -1.36$, 95% CI = [−0.59, −0.49]. Participants were also slightly more likely to believe that the hidden agent had intervened after negative outcomes in the neutral condition, although this effect was not significant, $t(254) = 1.62$, $p = .11$, $d = 0.10$, 95% CI = [−0.01, 0.12].

Computational modeling

Once again, we found that participants had significantly higher learning rates for positive outcomes than for negative outcomes, $t(254) = 3.57$, $p < .001$, $d = 0.22$, 95% CI = [0.00, 0.01]. In a further replication of our results from Experiment 1, we also found that learning rates were significantly lower for trials in which participants believed that the hidden agent intervened, compared with trials in which they believed that the hidden agent did not intervene, $t(252) = 17.55$, $p < .001$, $d = 1.10$, 95% CI = [0.04, 0.05] (Fig. 4b).