Supplementary Materials

Supplementary Study 1 Results

Table S1. Mean component scores by scenario condition. Standard errors are in parentheses.

	Tragic–Taboo	Tragic–Control	Taboo–Tragic	Taboo–Control
Factor1 (dispositional PE)	0.79 (0.13)	-0.80 (0.08)	0.91 (0.13)	-0.90 (0.08)
Factor 2 (prescriptive PE)	0.45 (0.18)	-0.52 (0.18)	-0.47 (0.18)	0.54 (0.18)
Factor 3 (mental imagery)	0.16 (0.20)	-0.48 (0.20)	0.43 (0.20)	-0.11 (0.20)
Factor 4 (descriptive PE)	-0.09 (0.21)	0.03 (0.21)	-0.01 (0.21)	0.07 (0.21)
Factor 5 (arousal)	0.06 (0.21)	-0.19 (0.21)	0.13 (0.21)	-0.002 (0.21)

Study 2 Model Specifications

Random effects parameters were chosen by first including: (1) by-subject and by-scenario random intercepts; (2) by-subject and by-scenario random slopes for all main effects; and (3) by-subject and by-scenario random slopes for significant interaction effects. Then, we removed random effects components that showed near-zero variance in an uncorrelated model until convergence could be achieved. Below we list maximal model specifications for illustrative purposes.

Condition-based Analyses. A linear mixed effects model was fit to predict PSC in the ToMN during the *initial segment*, including as fixed effects: Initial Condition (*tragic*, *taboo*), ROI (*DMPFC*, *RTPJ*, *LTPJ, PC*), and their interaction. A separate model was fit to predict PSC during the *final segment*, including as fixed effects: Initial Condition, Reframing Condition (*reframed*, *control*), ROI, PSC during the initial segment, and all interactions between Initial Condition, Reframing Condition, and ROI.

$PSC_{initial}$ ~ Initial Condition * ROI + (Initial Condition * ROI | subject) + (Initial Condition * ROI | scenario)

PSC_{final} ~ Initial Condition * Reframing Condition * ROI + $PSC_{initial}$ + (Initial Condition * Reframing Condition * ROI + $PSC_{initial}$ | subject) + (Initial Condition * Reframing Condition * ROI + $PSC_{initial}$ | scenario)

Correlation Analyses. Linear mixed effects models were fit to predict PSC during the *final segment*, including as fixed effects: one of the five components (dispositional prediction error, prescriptive prediction error, descriptive prediction error, mental imagery, and arousal), ROI, their interaction, and PSC during the initial segment.

PSC_{final} ~ component * ROI + $PSC_{initial}$ + (prediction error * ROI + $PSC_{initial}$ | subject) + (prediction error * ROI + $PSC_{initial}$ | scenario)

A linear mixed effects model was fit to predict moral judgments, including as fixed effects: PSC during the final segment, ROI, their interaction, and PSC during the initial segment.

how wrong ~ PSC_{final} * ROI + $PSC_{initial}$ + (PSC_{final} * ROI + $PSC_{initial}$ | subject) + (PSC_{final} * ROI + $PSC_{initial}$ | scenario)

Supplementary Study 2 Results

Table S2. Mean percent signal change in the Theory of Mind Network, by scenario condition. Standard errors are in parentheses.

	Initial Tragic	Initial Taboo	Tragic–Taboo	Tragic–Control	Taboo–Tragic	Taboo–Control
DMPFC	0.273 (0.061)	0.255 (0.061)	0.263 (0.049)	0.175 (0.048)	0.216 (0.048)	0.182 (0.048)
RTPJ	0.127 (0.039)	0.134 (0.039)	0.163 (0.036)	0.175 (0.032)	0.153 (0.036)	0.199 (0.034)
LTPJ	0.322 (0.043)	0.322 (0.043)	0.425 (0.035)	0.298 (0.035)	0.334 (0.035)	0.257 (0.035)
PC	0.169 (0.021)	0.179 (0.021)	0.236 (0.032)	0.114 (0.033)	0.183 (0.032)	0.077 (0.033)
ToMN	0.223 (0.034)	0.223 (0.034)	0.274 (0.016)	0.184 (0.017)	0.223 (0.016)	0.181 (0.017)

PSC for final segment. There was marginally greater ToMN activity for *tragic–taboo* vs. *taboo–tragic* scenarios (Estimate = 0.051, SE = 0.021, t(1657) = 2.469, p = 0.065), but no difference in ToMN activity for *tragic–control* vs. *taboo–control* scenarios (Estimate = 0.003, SE = 0.021, t(1664) = 0.130, p = 0.999).

There were no other significant interactions with ROI (3-way: Chisq(3) = 0.158, p = 0.984; Initial Condition x ROI: Chisq(3) = 3.467, p = 0.325).

Behavioral Component Score–ToM Activity Analysis. Activity in ToMN and in each ROI was not significantly associated with prescriptive prediction error, descriptive prediction error, mental imagery, or arousal.

Table S3. Relationships between component scores and Theory of Mind activity during the final segment. P-values for within-ROI effects were adjusted using the Holm-Bonferroni method.

	DMPFC	RTPJ	LTPJ	PC	ToMN
Factor1 (dispositional PE)	Estimate = 0.036, SE = 0.018, t(303.83) = 2.03, p = 0.087	Estimate = -0.007, SE = 0.014, t(20.22) = -0.53, p = 0.600	Estimate = 0.056, SE = 0.013, t(396.64) = 4.44, p < 0.0001	Estimate = 0.057, SE = 0.015, t(22.13) = 3.83, p = 0.003	Estimate = 0.038, SE = 0.007, t(1648) = 5.18, p < 0.0001
Factor 2 (prescriptive PE)	Estimate = 0.019, SE = 0.020, t(289.40) = 0.97, p = 0.960	Estimate = 0.013, SE = 0.013, t(17.57) = 1.01, p = 0.960	Estimate = 0.056, SE = 0.014, t(21.80) = 3.99, p = 0.028	Estimate = 0.003, SE = 0.015, t(399.4) = 0.17, p = 0.960	Estimate = 0.017, SE = 0.008, t(1351) = 2.13, p = 0.033
Factor 3 (mental imagery)	Estimate = -0.012, SE = 0.018, t(287.17) = -0.65, p = 1.000	Estimate = -0.002, SE = 0.012, t(413.33) = -0.20, p = 1.000	Estimate = -0.005, SE = 0.013, t(418.59) = -0.37, p = 1.000	Estimate = 0.0002, SE = 0.015, t(399.3) = 0.02, p = 1.000	Estimate = 0.003, SE = 0.008, t(1390) = 0.36, p = 0.719
Factor 4 (descriptive PE)	Estimate = -0.004, SE = 0.019, t(297.02) = -0.22, p = 1.000	Estimate = -0.007, SE = 0.012, t(423.77) = -0.56, p = 1.000	Estimate = -0.002, SE = 0.013, t(425.71) = -0.14, p = 1.000	Estimate = 0.021, SE = 0.015, t(393.78) = 1.42, p = 0.624	Estimate = -0.001, SE = 0.008, t(1289) = -0.11, p = 0.913
Factor 5 (arousal)	Estimate = -0.021, SE = 0.017, t(299.72) = -1.20, p = 0.700	Estimate = -0.012, SE = 0.012, t(426.32) = -1.07, p = 0.700	Estimate = -0.011, SE = 0.013, t(424.91) = -0.82, p = 0.700	Estimate = 0.002, SE = 0.014, t(420.2 = 0.13, p = 0.700	Estimate = -0.010, SE = 0.007, t(1607) = -1.36, p = 0.175