TESTING THE INTUITIVE RETRIBUTIVISM DUAL-PROCESS MODEL: ADDITIONAL DETAILS ON PLANNED CONTRAST ANALYSIS AND POWER ANALYSIS

1. Details on planned contrast analysis

To test our hypotheses, we performed a planned contrast analysis. We closely followed the procedure described by Schad et al. (2020). We first combined condition and motive (retributivism, deterrence, incapacitation) into one factor (ConditionxMotive) with six levels. We then expressed the nulls corresponding to our five hypotheses as contrasts of group means indexed by the levels of this factor. For example, the null corresponding to h1a can be expressed as:

$$1 \cdot \mu_{CR} + (-1) \cdot \mu_{CD} + 0 \cdot \mu_{CI} + 0 \cdot \mu_{TR} + 0 \cdot \mu_{TD} + 0 \cdot \mu_{TI} = 0 \qquad (1)$$

The μ_x are the mean rank-preferences scores of participants in group x. For example, μ_{CR} is the mean retributive rank-preference score in the control condition; μ_{TD} is the mean deterrence rank-preference score in the treatment condition; etc. In other words, then, (1) states that in the control condition, there will be no difference between the mean retributivism and mean deterrence rank-preference score.

Once expressed in this way, we extracted all the contrast coefficients and combined them in the following matrix:

$$\begin{pmatrix} 1 & 1 & 1 & 0 & 0 \\ -1 & 0 & 0 & 1 & 0 \\ 0 & -1 & 0 & 0 & 1 \\ 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & -1 \end{pmatrix}$$

Each column of this matrix contains the contrast coefficients of one null hypothesis. For example, comparison with (1) shows that the first column contains the coefficients corresponding to h1a. This matrix is then converted into what Schad et al. (2020) call a contrast matrix by applying the generalized matrix inverse. This contrast matrix has the correct format to specify the desired contrasts for a factor in R. The contrast matrix has full rank, so we can test all five contrasts in the same model.

To this end, we entered ConditionxMotive into a linear mixed-effects model predicting RPS (Bates et al., 2015). We added a random intercept for participant. In our preregistration, we had planned to add an additional random intercept for type of crime (blackmail, stolen property, arson, aggravated assault, murder); however, this model did not converge. Thus, we fit the model:

$$\text{RPS}_i = \beta_0 + \beta_1\text{ConditionxMotive}_{h1a,i} + \beta_2\text{ConditionxMotive}_{h1b,i} + \beta_3\text{ConditionxMotive}_{h2a,i} + \beta_4\text{ConditionxMotive}_{h2b,i} + \beta_5\text{ConditionxMotive}_{h2c,i} + u_{0i} + \varepsilon_i \,, \tag{2}$$

where i indexes the participant, $u_{0i} \sim N(0, \sigma^2_u)$, $\varepsilon_i \sim N(0, \sigma^2_\varepsilon)$. Moreover, h1a-h2c indicate the contrast being evaluated. Thus, estimating β_1 will provide a test of h1a; estimating β_2 will provide a test of h1b; etc.

2. Details on power analysis

We chose a simulation-based approach to power analysis (Brysbaert & Stevens, 2018), applied to the model in (2) with an added random intercept $u_{0j} \sim N(0, \sigma^2_v)$ for type of crime:

$$\text{RPS}_{ij} = \beta_0 + \beta_1\text{ConditionxMotive}_{h1a,ij} + \beta_2\text{ConditionxMotive}_{h1b,ij} + \beta_3\text{ConditionxMotive}_{h2a,ij} + \beta_4\text{ConditionxMotive}_{h2b,ij} + \beta_5\text{ConditionxMotive}_{h2c,ij} + u_{0i} + v_{0j} + \varepsilon_{ij}$$

The smallest mean difference between retributivism rank-preference score and either of the two utilitarian rank-preference scores reported by Keller et al. (2010 Exp. 2) was 3.18; we thus conservatively chose fixed effect sizes of $\beta = 2.0$ for the two contrasts corresponding to h1a and h1b. For each of the contrasts corresponding to the remaining three hypotheses, we estimated the smallest effect sizes of interest (Albers & Lakens, 2018) to be $\beta = 1.5$.

The power analysis was run using the simr package (Green & MacLeod, 2016, nsim = 2000). Using the estimates $\sigma^2_u = 1.0$, $\sigma^2_v = 1.0$ and $\sigma^2_\varepsilon = 4.0$, it indicated that in order to detect the fixed effects specified above at a level of significance of $\alpha = 0.01$ (Bonferroni corrected) with power of at least 90% (Chambers et al., 2019), a sample of size $n > 485$ participants would be required.

3. References

Albers, C., & Lakens, D. (2018). When power analyses based on pilot data are biased: Inaccurate effect size estimators and follow-up bias. *Journal of Experimental Social Psychology*, *74*, 187–195. https://doi.org/10.1016/j.jesp.2017.09.004

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, *67*(1). https://doi.org/10.18637/jss.v067.i01

Brysbaert, M., & Stevens, M. (2018). Power Analysis and Effect Size in Mixed Effects Models: A Tutorial. *Journal of Cognition*, *1*(1). https://doi.org/10.5334/joc.10

Chambers, C., Banks, G. C., Bishop, D., Bowman, S., Button, K., Crockett, M., Dienes, Z.,

Errington, T., Fischer, A., & Holcombe, A. O. (2019). Registered Reports. *OSF*. https://osf.io/8mpji/
Green, P., & MacLeod, C. J. (2016). SIMR : an R package for power analysis of generalized linear mixed models by simulation. *Methods in Ecology and Evolution*, *7*(4), 493–498. https://doi.org/10.1111/2041-210X.12504
Keller, L. B., Oswald, M. E., Stucki, I., & Gollwitzer, M. (2010). A Closer Look at an Eye for an Eye: Laypersons' Punishment Decisions Are Primarily Driven by Retributive Motives. *Social Justice Research*, *23*(2–3), 99–116. https://doi.org/10.1007/s11211-010-0113-4
Schad, D. J., Vasishth, S., Hohenstein, S., & Kliegl, R. (2020). How to capitalize on a priori contrasts in linear (mixed) models: A tutorial. *Journal of Memory and Language*, *110*, 104038. https://doi.org/10.1016/j.jml.2019.104038