# *Supplementary Information for:*

# Parsing information flow in speeded cognitive tasks: The associations of $g$ with perception and decision time

Emily A. Willoughby and James J. Lee
University of Minnesota Twin Cities
Department of Psychololgy

February 16, 2021

## 1 Participant and data exclusion criteria

### 1.1 Data filtering

We aimed to filter our data as little as possible in order to capture the full range of variability in response times and accuracy across our participants. Nevertheless, some minor filtering criteria were applied to ensure the reliability and validity of the data. Although it is standard to exclude very fast RTs between 100 ms and 200 ms (Whelan, 2008) for reaction time data as these likely do not reflect the process of interest, Luce (1986) argued that valid RTs have a minimum value of at least 100 ms, which can be considered the bare minimum of time needed for physiological processes such as stimulus perception and for motor responses. We have therefore excluded trials faster than 100 ms.

The best method of trimming slow responses has been a subject of debate in reaction time research. Ratcliff (1993) argued that eliminating RTs above an absolute point of cutoff would reduce power in cases where individual response variability was relevant to the effect of interest; in such cases, eliminating slow outliers based on each individual's distribution would better preserve real data. Given our lack of *a priori* assumptions of the effects of variability on our analyses of interest, particularly in modeling diffusion rate as a predictor of cognitive ability, we have opted for a conservative method of eliminating individual trial outliers that fall above five standard deviations above that individual's mean. No data transformations were applied to our RT data.

**Number comparison task** The number comparison task yielded a total of 116,400 individual trials across 776 participants prior to filtering. Two participants were removed from data analysis for having unusually slow mean response times after the application of data filtering criteria described below; these mean RTs of 1.86 s and 2.47 s were over 10 standard deviations above the sample mean, indicating that the processes under investigation were not being reliably assessed in these participants. Of the remaining 774 participants, none were removed for other reasons, yielding an initial 116,100 response trials prior to filtering. Only 5 individual trials were excluded for being below 0.100 s. A total of 2,371 incorrect trials were excluded from analysis. Slow outliers were trimmed at 5 standard deviations above each individual's reaction time mean, which excluded an additional 408 individual trials.

**Tone comparison task** For this task, 764 participants generated a total of 114,600 trials prior to filtering. Five participants were excluded from data analysis; four had too few correct RT trials per condition to participate in ANCOVA and other analyses, and one participant was excluded for having a total mean percent correct trials below chance (46.57% correct). No participant means were judged as too slow for inclusion, as even extremely slow means and large individual variances fell within the normal sample distribution. Participant exclusion yielded a total of 113,850 unfiltered trials for 759 participants. A total of 31 individual trials below 0.100 s and a total of 7,470 incorrect trials were excluded from further analyses. Slow outliers were trimmed at 5 standard deviations above each individual's reaction time mean, which excluded an additional 633 individual trials. These criteria netted a final sample of 105,716 valid and correct tone task trials across 759 individuals.

**ICAR-16 sample test** 773 participants took the short-form cognitive assessment. No participants were excluded based on unusual scores; for example, the one person who scored 0 out of 16 possible items was included because the time taken to complete the test was not unusually fast for this individual and a score of 0 was well within the normal distributional range of scores. No participants yielded incomplete data on the 16 items. However, several participants did not complete both the ICAR-16 and the tone or number comparison tasks; only participants who had both valid ICAR-16 and valid number comparison data were included in number comparison analyses of cognitive ability, and only those who had valid ICAR-16 and tone comparison data were included in tone comparison analyses. A total of 768 individuals had valid cognitive assessment scores and number comparison scores, 754 had valid cognitive assessment and tone comparison data, and 752 had valid cognitive assessment, number comparison, and tone comparison data. For measure reliability analyses (Supplementary Section 3), the full sample of available valid data for each measure is analyzed.

# 2 Experimental protocol

Experimenters were instructed to monitor a variety of potential sources of error. They verbally confirmed with all participants their age, the status of their vision as normal or normal-corrected, and their status as native English speakers. Participants were given both printed and verbal instructions for the tasks, and were required to verbally affirm that they had no questions or uncertainties about the task procedure before they began. Participants were also instructed to administer keypresses with their dominant hand, to use only the pointer and middle fingers, and to keep fingers on the keys at all times. Participants then answered a short computerized questionnaire to confirm their age and sex before beginning the experiment.

# 3 Reliability analyses

## 3.1 ICAR-16 sample test

**Confirmatory factor analysis** We used confirmatory factor analysis to test a four-factor model of the ICAR-16 composed of letter and number sequences, matrix reasoning, 3D rotation and verbal reasoning. This model was fit using the lavaan package in *R* with full-information maximum likelihood (FIML). Model fit was strong, with a TLI of .966 and RMSEA of .022 (90% CI: .012, .031). The full four-factor model fit the data significantly better than a single-factor solution ($\chi^2_6 = 275.85$, $p < .001$), and far better than a four-factor solution that did not allow covariances among the four latent factors ($\chi^2_6 = 293.54$, $p < .001$). As expected, the indicators all showed significant positive factor loadings, with standardized coefficients ranging from .301 to .652 (Supplementary Table S1).

Additionally, we observed significant positive correlations among all four latent factors (Supplementary Table S2), indicating that participants who showed high ability in one dimension were more likely to show high ability in the others as well. Taken together, these results are consistent with use of the ICAR-16 as a good short-form measure of cognitive ability, with the advantage of its short administration time outweighing its limitations in the context of this study.

## 3.2 Reaction time

### 3.2.1 Reliability within conditions

Since much of our theoretical claims rest on differential relationships of $g$ with the distance versus perceptual manipulations, it is important to demonstrate that distance and perceptual effects are reliable in order to mitigate the possibility that the differential correlations merely reflect an unreliable measure of perceptual difficulty. We conducted split-half reliability tests at each level of distance and perceptual difficulty using the *splithalf.r* function from *R*'s Multicon package. This function treats individual participant RT trials as items and finds the average of the randomly split-half correlation for a data frame of items. Each reported value was conducted with 1000 simulations, which indicates the number of split-half reliabilities to compute of which the mean is used as the best estimate. These results are summarized in Supplementary Table S3.

Table S1: Standardized factor loadings of each ICAR-16 item on its own latent factor in the current sample. All loadings are significant at $p < .001$.

| Latent Factor | Indicator | $\beta$ |
|---|---|---|
| Letter & Number | LN.33 | .481 |
| Letter & Number | LN.34 | .564 |
| Letter & Number | LN.58 | .534 |
| Letter & Number | LN.7 | .462 |
| Matrix Reasoning | MR.45 | .506 |
| Matrix Reasoning | MR.46 | .434 |
| Matrix Reasoning | MR.47 | .340 |
| Matrix Reasoning | MR.55 | .454 |
| 3D Rotation | R3D.3 | .615 |
| 3D Rotation | R3D.4 | .652 |
| 3D Rotation | R3D.6 | .583 |
| 3D Rotation | R3D.8 | .593 |
| Verbal Reasoning | VR.16 | .301 |
| Verbal Reasoning | VR.17 | .406 |
| Verbal Reasoning | VR.19 | .377 |
| Verbal Reasoning | VR.4 | .509 |

*Note:* R3D = Three-dimensional Rotation, LN = Letter And Number series, VR = Verbal Reasoning, MR = Matrix Reasoning.

Table S2: Latent factor correlations for the ICAR-16 in the current sample.

| Factor 1 | Factor 2 | Correlation | $p$-value |
|---|---|---|---|
| LN | MR | .534 | $< .001$ |
| LN | R3D | .468 | $< .001$ |
| LN | VR | .559 | $< .001$ |
| MR | R3D | .409 | $< .001$ |
| MR | VR | .595 | $< .001$ |
| R3D | VR | .558 | $< .001$ |

*Note:* Primary factor loadings for each item are shown in bold. R3D = Three-dimensional Rotation, LN = Letter And Number series, VR = Verbal Reasoning, MR = Matrix Reasoning.

Additionally, we conducted reliability tests of alpha and omega using *R*'s Psych package with maximum likelihood for the diffusion model parameters diffusion rate ($v$) and non-decision time ($T_{er}$) over the four levels of each manipulation (Supplementary Table S4). For factor extraction method, maximum likelihood was chosen over minimum residual because the latter generated Heywood cases, likely due to the small number of item types (levels of difficulty). The similarity of reliability estimates for diffusion parameters over different levels of difficulty suggests that they are behaving like indicators of a single factor.

#### 3.2.2 Reliability between conditions

In addition to within-condition reliabilities, we computed the reliabilities of RT slopes over levels of Distance and Contrast for both tasks. This was achieved by computing within-subjects slopes for RT over levels of each manipulation (equally counterbalanced across levels) across two halves of a participant's trials divided into even and odd trial numbers, and then correlating the two halves. For the number comparison task, even-trial slopes correlated with odd-trial slopes over levels of Distance at $r = .552$ and over levels of Contrast at $r = .481$ (both $p < .001$), which are not significantly different from one another (Fisher transformation $z = 1.9$; $p = .06$). For the tone comparison task, slopes over levels of Distance were significantly more reliable (Fisher's $z = 10.3$; $p < .001$) than slopes over levels of Loudness, with even-odd trial slopes correlating at $r = .67$ ($p < .001$) over levels of frequency distance, and at $r = .27$ ($p < .001$) over levels of tone loudness. This lower reliability over the tone task's perceptual manipulation is consistent with other analyses that suggest cautioned interpretations of effect comparisons between manipulations.

# 4 Analysis detail

## 4.1 EZ diffusion model

Diffusion parameters $v$ = diffusion rate, $T_{er}$ = non-decision time, and $a$ = boundary separation are calculated by adapting the EZ diffusion equations from Wagenmakers (2009). These parameters can be derived from equations that make use of distributional characteristics of RT data: Proportion of correct responses ($P_c$), the mean RT for correct responses ($MRT$), and variance of RT over correct responses ($VRT$).

The "EZ2" method was developed by Grasman, Wagenmakers & van der Maas (2009) to introduce improvements to the EZ model. EZ2 differs from EZ in that the starting point is allowed to vary and parameters can be constrained across conditions. It also can make use of the differential values of variance and means for all (correct and incorrect) trials and for correct trials only. EZ2 can therefore use the starting values obtained through EZ and this more detailed summary of RT moments to attempt to minimize the sum of the squared errors between observed values and those predicted by EZ2 to achieve better estimates of the parameters of interest across different conditions (Wagenmakers, 2009). EZ2 parameters for this project were computed with software package described in and developed by Grasman et al. (2009) (see also: raoul.socsci.uva.nl/EZ2).

**Limitations.** We acknowledge a number of limitations inherent in the application of the diffusion model to our data. First, the EZ/EZ2 methods are unable to compute parameters when a participant (or a group mean) performs with perfect accuracy, perfectly chance (.50 for a 2-choice task) accuracy, or perfect inaccuracy for a given condition. In such cases, Wagenmakers, van der Maas & Grasman (2007) recommend an edge correction equivalent to one half of an error dependent on number of trials $n$ for that condition.

However, in cases where error rates are low and the number of trials per condition is small, such an edge correction severely biases estimates of non-decision time. We applied a more stringent edge correction of .001 in our application of EZ to model individual differences. This is a more potent limitation for moments obtained in the number-comparison task, where error rates are very low. Several authors (Grasman et al., 2009; Lerche & Voss, 2018; Wagenmakers et al., 2007) note that for data sets with low error rates, up to 250 trials per condition are optimal for accurate parameter estimates. In the tradeoff between number of trials per subjects and total sample size, we prioritized sample size in order to increase the statistical power to falsify our additive-factors prediction of a null interaction between IQ and a given experimental manipulation.

It is interesting to note that despite these limitations in our data, mean diffusion rates and non-decision times over conditions yielded good reliability statistics (Supplementary Section 3) and that the ICAR-16 scores in our sample predict individual diffusion rates over conditions approximately as well as RTs on their own, particularly in the tone task. We take the results of our diffusion-model analyses as supporting but insufficient evidence of our main hypothesis.

### 4.2 Bayesian estimates for interaction effects

As stated in the main text, we computed Bayes factors in support of the null hypothesis for each effect reported in a table (Jeffreys, 1961) with methods implemented in the "BayesFactor" *R* package. For correlations and univariate regressions, we used the method of Bayesian $t$-statistic conversion described in Rouder, Speckman, Sun, Morey & Iverson (2009). For main effects reported from ANOVA designs, we used the $F$-statistic conversion described in Rouder, Morey, Speckman & Province (2012).

Though these methods work well for one-way and univariate models, Bayes factor estimation for interaction terms in ANCOVA is more complex. We opted to use Rouder, et al. (2012)'s *generalTestBF* method of estimating these effects, which models the Bayes factor in support of the null for each interaction effect by comparing the full model to the model with all effects except the one of interest. Since these interaction terms are estimated with a margin of error, their output is reported in full in Tables S5 (number task) and Table S6 (tone task).

## 5 Comparison of effects across both tasks

If the decision and perceptual components of both the number and tone comparison tasks represent global information-processing stages, it would logically follow that the effects can be compared in an aggregate model of both tasks.

This allows us to investigate the question of whether IQ interacts with Decisional difficulty (numerical distance + frequency distance), while simultaneously failing to interact with Perceptual difficulty (numeral contrast + tone loudness).

A total of 752 participants had valid data for the ICAR-16 and all conditions for both number and tone comparison tasks. Individual mean reaction times correlated strongly across tasks ($r = .594$, $p < .001$), as did individual diffusion rate $v$ ($r = .491$, $p < .001$). Non-decision time $T_{er}$ was more variable but nevertheless correlated across tasks ($r = .304$, $p < .001$). Despite high accuracy in the number comparison task, error rates between tasks also correlated modestly ($r = .155$, $p < .001$).

We evaluated the interaction of IQ with both types of manipulation through analysis of covariance, comparison of regression slopes, and linear-mixed modeling on combined data of both tasks. Mean RTs were collapsed first for manipulation levels for each task and then aggregated across tasks.

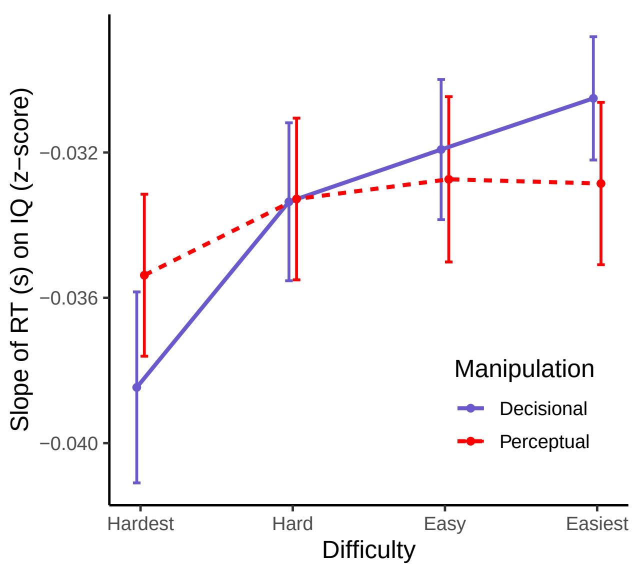


Figure S1: Regression coefficient (slope) of RT (s) on IQ across Decisional and Perceptual difficulty for both tasks. Error bars represent ±1 standard error.

## 5.1 ANCOVA and regression slopes

Aggregate ANCOVA reveals a small but reliable interaction between IQ and the Decisional manipulation ($F_{3,2250} = 7.32$, $p < .001$). By contrast, no evidence was observed for a simultaneous significant interaction between IQ and the Perceptual manipulation ($F_{3,2250} = 1.51$, $p = .214$). Task revealed a powerful main effect of $F_{1,750} = 162.81$ ($p < .001$) on RT, along with nominally significant interactions between Decisional and Perceptual manipulations ($F_{9,6750} = 2.55$, $p = .011$) and IQ × Perceptual × Decisional levels ($F_{9,6750} = 2.10$, $p = .037$). No significant evidence for higher-order interactions of IQ with manipulation and Task were observed. These effects and their associated statistics are shown in Table S7.

A comparison of regression slopes of RT (s) on IQ (standardized) with both tasks aggregated reveals strong visual evidence in favor of our hypothesis that IQ interacts with a Decisional, but not Perceptual, manipulation, regardless of the sensory modality of the stimuli (Supplementary Figure S1). With IQ scores standardized across the sample, each level of Distance is associated with unstandardized beta coefficients (SE) of −0.038 (0.003) for the hardest level, −0.033 (0.002) for the next-hardest level, −0.032 (0.002) for the easier level, and −0.031 (0.002) for the easiest level. This equates to an advantage conferred by each standard deviation gain in IQ to a reduction in mean RT of about 39 ms for the hardest Decisional level, with a reduction of only 31 ms for the easiest level. Over levels of Perceptual difficulty, the hardest condition conferred an RT reduction per each SD of IQ of 35 ms (SE = 2 ms), which is scarcely different from the advantage conferred at the easiest Perceptual level (33 ms, SE = 2 ms).

## 5.2 Linear mixed-modeling

A similar pattern of effects results from linear mixed-modeling, in which we include task as a fixed effect term and random effects terms for block (1–3) and stimulus digit/tone frequency (1–4 and 6–9), as well as interactions for IQ with Task, IQ with Decisional difficulty × Task, and IQ with Perceptual difficulty × Task. This results in strong main effects of both types of manipulation, with no evidence for a Decisional × Perceptual interaction outside of Task. Crucially, this model produced clear evidence of an interaction of IQ with Decisional difficulty ($F = 66.00$, $p < .001$) in the simultaneous absence of evidence for an interaction of IQ with Perceptual difficulty ($F = 0.85$, $p = .465$). IQ also significantly interacted with Task and played a role in several higher-order interactions with Task and both manipulations. These effects and their associated statistics are shown in Supplementary Table S8.

To visualize the results of this LMM, we used the *sjPlot* package in *R* to plot regression models for both types of manipulation as ordered variables. This enables the visual inspection of the magnitude of difference in slope of RT on IQ between Decisional and Perceptual manipulations (Supplementary Figure S2). While evidence for an IQ × Decisional interaction is strong across tasks, it remains unclear if the numeral contrast and tone loudness perceptual manipulations are truly analogous. For example, the three-way interaction between IQ, Perceptual and Task is significant in the linear mixed model ($p < .001$) but not in ANCOVA ($p = .307$), thus limiting the interpretability of these analyses across both tasks.

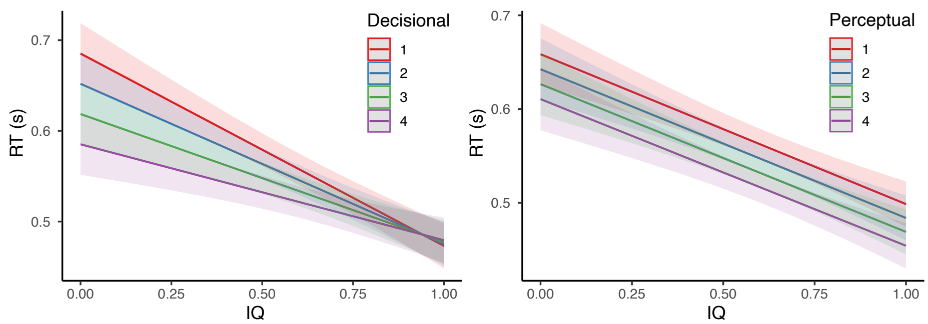


Figure S2: Regression of RT (s) on IQ for results obtained through LMM over both tasks in aggregate. Slopes show the effect of each manipulation as an ordered variable; i.e., "hardest" level of each manipulation = 1, "hard" = 2, "easy" = 3, and "easiest" = 4. Bands around each line represent 95% confidence intervals.

# References

Grasman, R. P. P., Wagenmakers, E.-J., & van der Maas, H. L. J. (2009). On the mean and variance of response times under the diffusion model with an application to parameter estimation. *Journal of Mathematical Psychology*, *53*(2), 55–68. doi:10.1016/j.jmp.2009.01.006

Jeffreys, H. (1961). *The theory of probability*. Oxford, UK: Oxford University Press.

Lerche, V., & Voss, A. (2018). Speed-accuracy manipulations and diffusion modeling: Lack of discriminant validity of the manipulation or of the parameter estimates? *Behavior Research Methods*, *50*, 2568–2585. doi:10.3758/s13428-018-1034-7

Luce, R. D. (1986). *Response times: Their role in inferring elementary mental organization*. New York, NY: Oxford University Press.

Ratcliff, R. (1993). Methods for dealing with reaction time outliers. *Psychological Bulletin*, *114*(3), 510–532. doi:10.1037/0033-2909.114.3.510

Rouder, J. N., Morey, R. D., Speckman, P. L., & Province, J. M. (2012). Default Bayes factors for ANOVA designs. *Journal of Mathematical Psychology*, *56*(5), 356–374. doi:10.1016/j.jmp.2012.08.001

Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D., & Iverson, G. (2009). Bayesian $t$ tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin & Review*, *16*, 225–237. doi:10.3758/PBR.16.2.225

Wagenmakers, E.-J. (2009). Methodological and empirical developments for the Ratcliff diffusion model of response times and accuracy. *European Journal of Cognitive Psychology*, *21*(5), 641–671. doi:10.1080/09541440802205067

Wagenmakers, E.-J., van der Maas, H. L. J., & Grasman, R. P. P. (2007). An EZ-diffusion model for response time and accuracy. *Psychonomic Bulletin and Review*, *14*(1), 3–22. doi:10.3758/BF03194023

Whelan, R. (2008). Effective analysis of reaction time data. *Psychological Record*, *58*(3), 475–482. doi:10.1007/BF03395630

Table S3: Reliability comparisons of reaction time parameters across levels of Distance and Contrast or Loudness for both number and tone comparison tasks

| Number comparison task | | | | | Tone comparison task | | | |
|---|---|---|---|---|---|---|---|---|
| *Mean reaction time* | | | | | *Mean reaction time* | | | |
| | Mean split-half $r$ | Mean reliability | $SD$ reliability | | | Mean split-half $r$ | Mean reliability | $SD$ reliability |
| *Distance* | | | | | *Distance* | | | |
| Hardest | 0.87 | 0.93 | 0.04 | | Hardest | 0.90 | 0.95 | 0.11 |
| Hard | 0.87 | 0.93 | 0.04 | | Hard | 0.89 | 0.94 | 0.11 |
| Easy | 0.88 | 0.94 | 0.05 | | Easy | 0.90 | 0.95 | 0.14 |
| Easiest | 0.89 | 0.94 | 0.05 | | Easiest | 0.91 | 0.95 | 0.14 |
| *Contrast* | | | | | *Loudness* | | | |
| Hardest | 0.88 | 0.94 | 0.04 | | Hardest | 0.89 | 0.94 | 0.11 |
| Hard | 0.87 | 0.93 | 0.04 | | Hard | 0.89 | 0.94 | 0.13 |
| Easy | 0.87 | 0.93 | 0.05 | | Easy | 0.90 | 0.95 | 0.11 |
| Easiest | 0.86 | 0.93 | 0.04 | | Easiest | 0.88 | 0.94 | 0.12 |

*Mean split-half r* = average of all split-half correlations, *Mean reliability* = average of all split-half reliabilities, *SD reliability* = standard deviation of all split-half reliabilities. Diffusion model parameter reliabilities are computed over the four levels of each manipulation. $\omega_h$ = omega hierarchical, $\omega_t$ = omega total.

Table S4: Reliability comparisons of diffusion model parameters across levels of Distance and Contrast or Loudness (labeled *Perceptual*) for both number and tone comparison tasks

| | $\omega$ Number | $\omega$ Tone |
|---|---|---|
| *Distance* | | |
| $v$ | 0.83 | 0.88 |
| $T_{er}$ | 0.77 | 0.89 |
| *Perceptual* | | |
| $v$ | 0.85 | 0.91 |
| $T_{er}$ | 0.87 | 0.87 |

*Note:* $\omega$ = omega hierarchical.

Table S5: Number task: BF01 estimates for interaction effects from *generalTestBF*. Each Bayes factor represents the effect when the term is omitted from the full model.

| Omitted effect | BF01 | Error |
|---|---|---|
| Contrast:Distance | 4868.33 | $\pm$9.93% |
| IQ:Contrast | 408.73 | $\pm$29.37% |
| IQ:Distance | $3.07 \times 10^{-6}$ | $\pm$16.21% |
| IQ:Contrast:Distance | $1.85 \times 10^{9}$ | $\pm$10.02% |

Table S6: Tone task: BF01 estimates for interaction effects from *generalTestBF*. Each Bayes factor represents the effect when the term is omitted from the full model.

| Omitted effect | BF01 | Error |
|---|---|---|
| Loudness:Distance | 4332.9 | $\pm$15.73% |
| IQ:Loudness | 1483.5 | $\pm$12.44% |
| IQ:Distance | 0.143 | $\pm$11.71% |
| IQ:Loudness:Distance | $7.45 \times 10^{8}$ | $\pm$11.85% |

Table S7: ANCOVA table of effects for both tasks aggregated over Decisional difficulty (numerical distance + tone frequency) and Perceptual difficulty (numeral contrast + tone loudness).

| Variables | $DF_n$ | $DF_d$ | $F$-value | $p$-value | BF01 |
|---|---|---|---|---|---|
| Perceptual | 3 | 2250 | 306.34 | $< .001$ | $2.9 \times 10^{-170}$ |
| Decisional | 3 | 2250 | 1194.15 | $< .001$ | $\sim 0$ |
| Task | 1 | 750 | 162.81 | $< .001$ | $8.3 \times 10^{-33}$ |
| **IQ:Perceptual** | **3** | **2250** | **1.51** | **.214** | **498.7** |
| **IQ:Decisional** | **3** | **2250** | **7.32** | **< .001** | **0.014** |
| IQ:Task | 1 | 750 | 13.17 | $< .001$ | 0.03 |
| Perceptual:Decisional | 9 | 6750 | 2.55 | .011 | 7274 |
| IQ:Perceptual:Decisional | 9 | 6750 | 2.10 | .037 | $2.8 \times 10^{8}$ |
| IQ:Perceptual:Task | 3 | 2250 | 1.20 | .307 | 627.4 |
| IQ:Decisional:Task | 3 | 2250 | 0.62 | .521 | 629.6 |
| IQ:Perceptual:Decisional:Task | 9 | 6750 | 0.88 | .534 | $2.9 \times 10^{9}$ |

*Note:* $P$-values reported from repeated-measures covariance analyses reflect Greenhouse-Geisser sphericity corrections. Column BF01 reflects the Bayes factor in support of the null hypothesis. Effects and associated statistics relevant to core hypotheses are bolded.

Table S8: ANOVA summary of linear mixed-model effects and interactions for both tasks aggregated over Decisional difficulty (numerical distance + tone frequency) and Perceptual difficulty (numeral contrast + tone loudness). Random effects terms are included for subject ID, block (1–3), and frequency/digit of each task's Decisional manipulation (1–4 and 6–9).

| Variables | Groups | $F$-value | $p$-value | Estimate (SE) | BF01 |
|---|---|---|---|---|---|
| IQ | 1 | 56.80 | $< .001$ | $-0.234$ (0.028) | $1.7 \times 10^{-11}$ |
| Perceptual | 4 | 50.93 | $< .001$ | $-0.010$ (0.004) | $2.2 \times 10^{-29}$ |
| Decisional | 4 | 154.82 | $< .001$ | $-0.027$ (0.004) | $2.6 \times 10^{-90}$ |
| Task | 2 | 1318.72 | $< .001$ | 0.127 (.003) | $2.9 \times 10^{-204}$ |
| **IQ:Perceptual** | **4** | **0.85** | **.465** | **$-0.005$ (0.006)** | **$7.4 \times 10^{7}$** |
| **IQ:Decisional** | **4** | **66.00** | **$< .001$** | **0.029 (0.006)** | **0.0016** |
| IQ:Task | 2 | 41.87 | $< .001$ | $-0.030$ (0.011) | $2.2 \times 10^{-69}$ |
| Percep:Decis | 16 | 1.37 | .197 | $-0.002$ (0.001) | $2.2 \times 10^{15}$ |
| IQ:Percep:Decis | 16 | 1.47 | .154 | 0.003 (0.002) | $9.2 \times 10^{15}$ |
| IQ:Percep:Task | 8 | 30.78 | $< .001$ | 0.017 (0.003) | $4.9 \times 10^{-39}$ |
| IQ:Decis:Task | 8 | 152.87 | $< .001$ | $-0.036$ (0.003) | $3.8 \times 10^{-205}$ |
| IQ:Per:Decis:Task | 32 | 1.05 | .398 | $-0.001$ (0.001) | $3.2 \times 10^{13}$ |

*Note*: Column BF01 reflects the Bayes factor in support of the null hypothesis. Effects and associated statistics relevant to core hypotheses are bolded. "Estimate" refers to the slope of RT (s) on IQ for levels of each manipulation as ordered variables; i.e., "hardest" level of each manipulation = 1, "hard" = 2, "easy" = 3, and "easiest" = 4. Reported Bayes factors were computed on each effect's $F$-value against the default prior.